QoE-aware Media Streaming in Technology and Cost Heterogeneous Networks

Ali ParandehGheibi, Asuman Ozdaglar and Muriel Médard

arXiv:1203.3258v1 [cs.SY] 15 Mar 2012

Abstract

We present a framework for studying the problem of media streaming in technology and cost heterogeneous environments. We first address the problem of efficient streaming in a technology-heterogeneous setting. We employ random linear network coding to simplify the packet selection strategies and alleviate issues such as duplicate packet reception. Then, we study the problem of media streaming from multiple cost-heterogeneous access networks. Our objective is to characterize analytically the trade-off between access cost and user experience. We model the Quality of user Experience (QoE) as the probability of interruption in playback as well as the initial waiting time. We design and characterize various control policies, and formulate the optimal control problem using a Markov Decision Process (MDP) with a probabilistic constraint. We present a characterization of the optimal policy using the Hamilton-Jacobi-Bellman (HJB) equation. For a fluid approximation model, we provide an exact and explicit characterization of a threshold policy and prove its optimality using the HJB equation.

Our simulation results show that under properly designed control policy, the existence of alternative access technology as a complement for a primary access network can significantly improve the user experience without any bandwidth over-provisioning.

Index Terms

Media streaming, Quality of Experience (QoE), Heterogeneous networking, Network association policy, Network coding.

I. INTRODUCTION

Media streaming is fast becoming the dominant application on the Internet [1]. The popularity of available online content has been accompanied by the growing usage of wireless handheld devices as the preferred means of media access. The predictions by Cisco Visual Networking Index [2] show that by 2015, the various forms of video (TV, VoD, Internet Video, and P2P) will exceed 90 percent of global consumer traffic, and 66 percent of total mobile traffic. In order to cope with the demand, the wireless service providers generally build new infrastructure. Another approach that seems to be gaining momentum in the industry is offloading mobile data traffic onto another network through dual-mode such as additional WiFi interfaces [3]. This approach requires the wireless devices to operate seamlessly in an environment with heterogeneous access methods (WiFi, 3G and 4G) and different access costs (cf. Figure 1). For example, accessing public WiFi networks is free but unreliable, while there are additional charges associated with more reliable 3G or 4G data networks. The goal of this work is to design network access policies that minimize the access cost in such heterogeneous environments, while guaranteeing an acceptable level of quality of user experience.

Fig. 1.   Media streaming from heterogeneous servers.

In particular, we focus on media streaming applications which are intrinsically delay-sensitive. Hence, they need to be managed differently from the traditional less delay-sensitive applications such as Web, Email and file downloads. Most of the current approaches for providing a reasonable Quality of Service (QoS) for streaming applications are based on resource over-provisioning,

without considering the transient behavior of the service provided to such applications. In this work, we pay special attention to communication and control techniques that are specifically designed with Quality of user Experience (QoE) in mind. The goal, on one hand, is to make the optimal use of the limited and possibly unreliable resources to provide a seamless experience for the user. On the other hand, we would like to provide a tool for the service providers to improve their service delivery (specifically for delay-sensitive applications) in the most economical way. Our contributions are summarized in the following.

We first address the problem of efficient streaming in technology-heterogeneous settings, where a user receives a stream over multiple paths from different servers. Each sever can be a wireless access point or another peer operating as a server. We consider a model that the communication link between the receiver and each server is unreliable, and hence, it takes a random period of time for each packet to arrive at the receiver from the time that the packet is requested from a particular server. One of the major difficulties with such multi-server systems is the packet tracking and duplicate packet reception problem, i.e., the receiver need to keep track of the index of the packets it is requesting from each server to avoid requesting duplicate packets. Since the requested information is delay sensitive, if a requested packet does not arrive within some time interval, the receiver need to request the packet from another server. This may eventually result in receiving duplicate packets and waste of resources. We address this issue and discuss that using random linear network coding (RLNC)[4] across packets within each block of the media file we can alleviate this issue. This technique assures us that, with high probability, no redundant information will be delivered to the receiver.

We would like to emphasize that one of the critical roles of network coding techniques in this work, other than efficient and seamless streaming, is to simplify greatly the communication models, so that we can focus on end-user metrics and trade-offs. For example, if each server can effectively transmit packets according to an independent Poisson process, using RLNC we can merge these processes into one Poisson process of sum rate. Hence, the system model boils down to a single-server single-receiver system.

We then study the problem of media streaming in a cost-heterogeneous environment. We consider a system wherein network coding is used to ensure that packet identities can be ignored, and packets may potentially be obtained from two classes of servers with different rates of transmission. The wireless channel is unreliable, and we assume that each server can deliver

packets according to a Poisson process with a known rate. Further, the costs of accessing the two servers are different; for simplicity we assume that one of the servers is free. Thus, *our goal is to develop an algorithm that switches between the free and the costly servers in order to satisfy the desired user experience at the lowest cost.*

The user experience metrics that we consider in this work are the initial buffering *delay* before the media playback, and the probability of experiencing an interruption throughout media playback. Interruption probability captures the reliability of media playback. Such metrics best capture the user experience for most of media streaming applications e.g. Internet video, TV, Video on Demand (VoD), where the user may tolerate some initial delay, but expects a smooth sequential playback. In [5], we characterized the optimal trade-off between these metrics for a single-server single-receiver system. In particular, we established the following relation

$$\text{Probability of interruption} = e^{-I(R)\cdot(\text{initial buffering})}, \tag{1}$$

where $I(R)$ is the *interruption exponent* or reliability function, which depends the arrival rate R of the stream. This result is analogous to information theoretic error exponent results relating the error probability of a code to the block length of that code.

In a cost-heterogeneous system, the user experience such as initial waiting time may be improved by simultaneously accessing free and costly access methods. This adds another dimension to the problem the end-user is facing. Certain levels of user satisfaction can only be achieved by paying a premium for extra resource availability. Figure 2 illustrates a conceptual three-dimensional cost-delay-reliability trade-off curve.

The objective of this paper is to understand the trade-off between initial waiting time, and the usage cost for attaining a target probability of interruption, and design control policies to achieve the optimal trade-off curve. We study several classes of server selection policies. Using the QoE trade-offs for a single-server system, we obtain a lower bound on the cost of offline policies that do not observe the trajectory of packets received. We show that such policies have a threshold form in terms of the time of association with the costly server. Using the offline algorithm as a starting point, we develop an *online* algorithm with lower cost that has a threshold form – both free and costly servers are used until the queue length reaches a threshold, followed by only free server usage. We then develop an online risky algorithm in which the risk of interruption is spread out across the trajectory. Here, only the free server is used whenever the
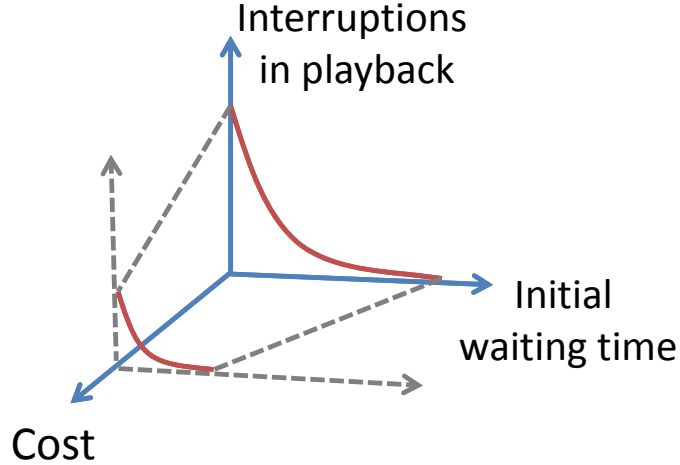
Fig. 2. Trade-off between the achievable QoE metrics and cost of communication.

queue length is above a certain threshold, while both servers are used when the queue length is below the threshold. The threshold is designed as a function of the initial buffering and the desired interruption probability. We numerically compare the performance of the proposed control policies. Our simulation results show online risky algorithm performs the best. Moreover, we observe that the existence of costly networks as a complement for unreliable but cheaper networks can significantly improve the user experience without incurring a significant cost.

We formulate the problem of finding the optimal network association policy as a Markov Decision Process (MDP) with a probabilistic constraint. Similarly to the Bellman equation proposed by Chen [6], for a discrete-time MDP with probabilistic constraints, we write the Hamilton-Jacobi-Bellman (HJB) equation for our continuous-time problem by proper state space expansion. The HJB equation is instrumental in optimality verification of a particular control policy for which the expected cost is explicitly characterized. However, due to discontinuity of the queue-length process for the Poisson arrival model, a closed-form characterizations of our proposed policies are not available for the verification of the HJB equation. Therefore, we consider a fluid approximation model, where the arrival process is modeled as a controlled Brownian motion with a drift. In this case, we provide and exact and explicit characterization of a threshold policy that satisfies the QoE constraints. We show that the expected cost corresponding to this threshold policy indeed is the solution of the corresponding HJB equation, thus proving

the optimality of such policies.

A. Related Work

The set of works related to this thesis spans several distinct areas of the literature. One of the major difficulties in the literature is the notion of *delay*, which greatly varies across different applications and time scales at which the system is modeled. The role of delay-related metrics has been extensive in the literature on Network Optimization and Control. Neely [7], [8] employs Lyapunov optimization techniques to study the delay analysis of stochastic networks, and its trade-offs with other utility functions. Other related works such as [9], [10] take the flow-based optimization approach, also known as Network Utility Maximization (NUM), to maximize the *delay-related* utility of the users. Closer to our work is the one by Hou and Kumar [11] that considers per-packet delay constraints and successful delivery ratio as user experience metrics. Such flow-based approaches are essentially operating at the steady state of the system, and fail to capture the end-user experience for delay-sensitive applications of interest.

Media streaming, particularly in the area of P2P networks, has attracted significant recent interest. For example, works such as [12], [13], [14], [15] develop analytical models on the trade-off between the steady state probability of missing a block and buffer size under different block selection policies. Unlike our model, they consider live streaming, e.g. video conferencing, with deterministic channels. However, we focus on content that is at least partially cached at multiple locations, and must be streamed over one or more unreliable channels. Further, our analysis is on transient effects—we are interested in the first time that media playback is interrupted as a function of the initial amount of buffering. Also related to our work is [16], which considers two possible wireless access methods (WiFi and UMTS) for file delivery, assuming particular throughput models for each access method. In contrast to this work, packet arrivals are stochastic in our model, and our streaming application requires hard constraints on quality of user experience.

Another body of related work is the literature on constrained Markov decision processes. There are two main approaches to these problems. Altman [17], Piunovskiy [18], [19] and Feinberg and Shwartz [20] take a convex analytic approach leading to linear programs for obtaining the optimal policies. On the other hand, Chen [6], Chen and Blankenship [21], Piunovskiy [22] use the more natural and straightforward Dynamic Programming approach to characterize all

optimal policies. These works mainly focus on different variations of the discrete-time Markov decision processes. In this work, we take the dynamic programming approach for the control of a continuous-time Markovian process. Further, we employ stochastic calculus techniques used in treatment of stochastic control problems [23] to properly characterize the optimal control policies.

The rest of this paper is organized as follows. In Section II, we discuss using network coding techniques to guarantee efficient streaming in a technology heterogeneous system. Section III describes the system model and QoE metrics for a media streaming scenario from cost-heterogeneous servers. In Section IV, we present and compare several server association policies. The dynamic programming approach for characterization of the optimal control policy is discussed in Section V. We present the fluid approximation model and establish the optimality of an online threshold policy in Section VI. Finally, Section VII provides a summary of the contributions of this paper with pointers for potential extensions in the future.

II. NETWORK CODING FOR EFFICIENT STREAMING IN TECHNOLOGY-HETEROGENEOUS SYSTEMS

In this part, we study the problem of streaming a media file from multiple servers to a single receiver over unreliable communication channels. Each of the servers could be a wireless access point, base station, another peer, or any combination of the above. Such servers may operate under different protocols in different ranges of the spectrum such as WiFi (IEEE 802.11), WiMAX (IEEE 802.16), HSPA, EvDo, LTE, etc. We refer to such system as a technology-heterogeneous multi-server system. In this setup, the receiver can request different pieces of the media file from different servers. Requesting packets form each server may cause delays due to channel uncertainty. However, requesting one packet from multiple servers introduces the need to keep track of packets, and the duplicate packet reception problem. In this section, we discuss methods that enable efficient streaming across different paths and network interfaces. This greatly simplifies the model when analyzing such systems.

In order to resolve such issues of the multi-server and technology-heterogeneous systems, let us take a closer look at the process of media streaming across different layers. Media files are divided into blocks of relatively large size, each consisting of several frames. The video coding is such that all the frames in the block need to be available before any frames can be played.

Blocks are requested in sequence by the playback application from the user-end. The server (or other peers) packetize the requested block and transmit them to the user as in Figure 3.
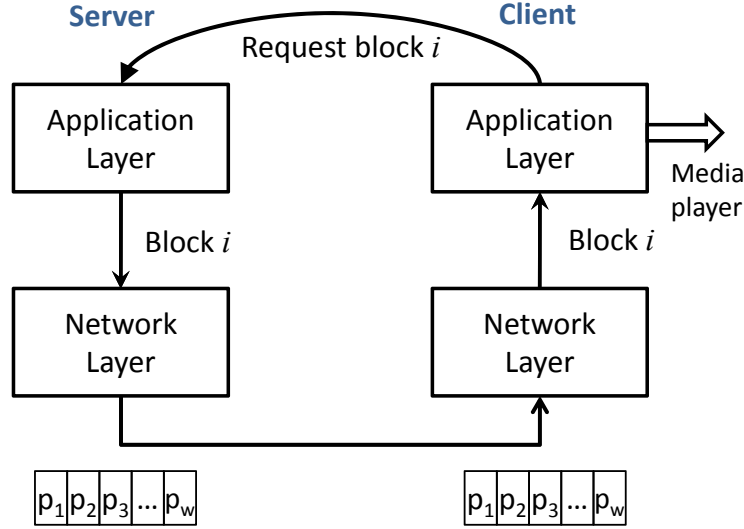

Fig. 3. The media player (application layer) requires complete blocks. At the network layer each block is divided into packets and delivered.

Now consider the scenario, illustrated in Figure 4, where there are multiple paths to reach a particular server. Each of these paths could pass through different network infrastructures. For example, in Figure 4, one of the paths is using the WiFi access point, while the other one is formed by the LTE (Long Term Evolution) network.

The conventional approach in exploiting the path diversity in such scenarios is scheduling each packet to be delivered over one of the available paths. For instance, odd-numbered packets are assigned to path 1, and even-numbered packets are assigned to path 2. This approach requires a deterministic and reliable setting, where each path is lossless and the capacity and end-to-end delay of each path is known. However, the wireless medium is intrinsically unreliable and time varying. Moreover, flow dynamics in other parts of the network may result in congestion on a particular path. Therefore, the slowest or most unreliable path becomes the bottleneck. In order to compensate for that, the scheduler may add some redundancy by sending the same packet over multiple paths, which results in duplicate packet reception and loss of performance. There is a significant effort to use proper scheduling and control mechanisms to reduce these problems. For more information on this approach, generally known as MultiPath TCP (MPTCP), please
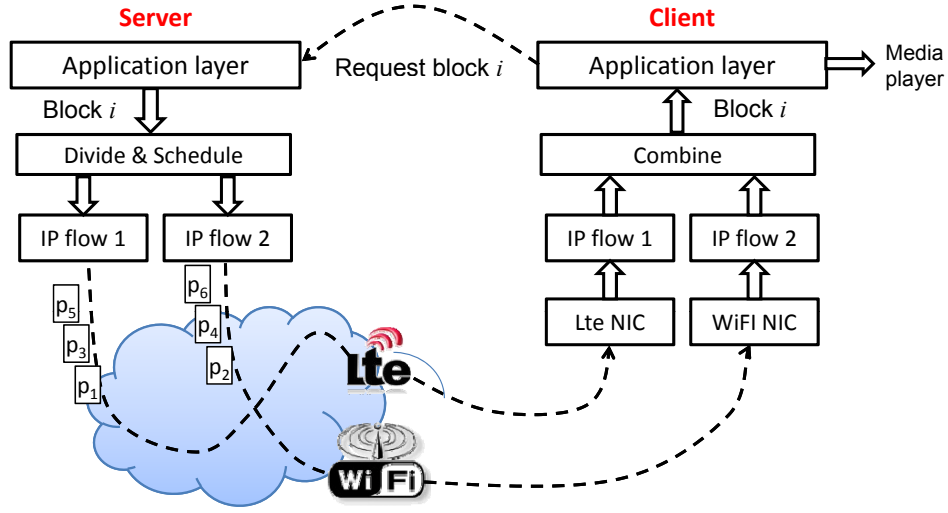
Fig. 4. Streaming over multiple paths/interfaces.

refer to the works by Wischik *et al.* [24], [25], and IETF working draft on MPTCP [26].
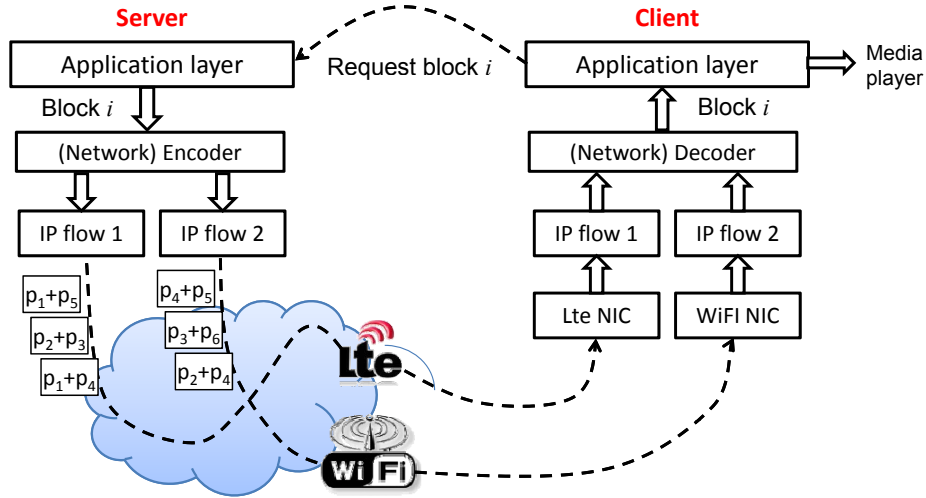
Fig. 5. Streaming over multiple paths/interfaces using Network Coding.

We propose random linear network coding (RLNC) to alleviate the duplicate packet reception problem. Figure 5 illustrates an example. Here, instead of requesting a particular packet in block i, the receiver simply requests a random linear combination of all the packets in block i. The coefficients of each combination are chosen uniformly at random from a Galois field of size

q. The coded packets delivered to the receiver can be thought of as linear equations, where the unknowns are the original packets in block i. Block i can be fully recovered by solving a system of linear equations if it is full rank. Note that we can embed the coding coefficients in the header of each coded packet so that the receiver can form the system of linear equations. For more implementation details, please refer to Sundararajan *et al.* [27]. It can be shown that if the field size q is large enough, the received linear equations are linearly independent with very high probability [4]. Therefore, for recovering a block of W packets, it is sufficient to receive W linearly independent coded packets from different peers. Each received coded packet is likely to be independent of previous ones with probability $1 - \delta(q)$, where $\delta(q) \to 0$ as $q \to \infty$.

By removing the notion of *unique identity* assigned to each packet, network coding allows significant simplification of the scheduling and flow control tasks at the server. For instance, if one of the paths get congested or drops a few of the packets, the server may complete the block transfer by sending more coded packets over the other paths. Hence, the sender may perform TCP-like flow management and congestion control on each of the paths *independently*. Therefore, network coding provides a mean to homogenize a technology-heterogeneous system as if the receiver only has single interface.

Note that such random linear coding does not introduce additional decoding delay for each block, since the frames in a block can only be played out when the whole block is received. So there is no difference in delay whether the end-user received W uncoded packets of the block or W independent coded packets that can then be decoded.

In the following, we discuss the conditions under which we can convert a technology-heterogeneous multi-server system to a single-path single-server system. Consider a single user receiving a media file from various servers it is connected to. Assume that the media file is divided into blocks of W packets. Each server sends random linear combinations of the packets within the current block to the receiver. We assume that the linear combination coefficients are selected from a Galois field of large enough size, so that no redundant packet is delivered to the receiver. Moreover, we assume that the block size W is small compared to the total length of the file, but large enough to ignore the boundary effects of moving from one block to the next. For simplicity of the analysis, we assume in this work that the arrival process of packets from each server is a Poisson process. Since, network coding allows for independent flow control on each path, we may assume that the arrival process from each server is independent of other arrival processes.

Moreover, since we can assume no redundant packet is delivered from different peers, we can combine the arrival processes into one Poisson process of the sum-rate R. Thus, our simplified model is just a single-server-single-receiver system.

We summarize the above discussions into the following Assumption, which is the key for development of the analytical results in the subsequent parts.

Assumption 1. Consider one or more servers streaming a single media file to a single client over m independent path using random linear network coding. The packet delivery process over path k is modeled as a Poisson process of rate R_k. The effective packet delivery process observed at the receiver is a Poisson process of rate $R = \sum_{k=1}^{m} R_k$.

Note that, for simplicity of the system model and analysis, we are neglecting a few complexities associated with network coding approach such as the effect of field size, feedback imperfections, and other uncertainties

Assumption 1 provides the necessary tool for analyzing technology-heterogeneous multi-server systems as a single-server system. For instance, we can apply most of the results of [5] on fundamental delay-interruption trade-offs. This is essential for tractability of the analysis of cost-heterogeneous systems, which is the focus of the subsequent part.

III. SYSTEM MODEL AND QoE METRICS

We consider a media streaming system as follows. A single user is receiving a media stream of infinite size, from various servers or access points. The receiver first buffers D packets from the beginning of the file, and then starts the playback at unit rate.

We assume that time is continuous, and the arrival process of packets from each server is a Poisson process independent of other arrival processes. Further, we assume that each server sends random linear combination of the packets in the source file. Therefore, by discussions of Section II, no redundant packet is delivered from different servers. Therefore, we can combine the arrival processes of any subset of the servers into one Poisson process of rate equal to the sum of the rates from the corresponding servers (cf. Assumption 1).

There are two types of servers in the system: free [1] servers and the costly ones. There is no

---

[1]The contributions of this work still hold if both servers are costly with different access costs. Here, we normalize the access cost of the cheaper server to zero.

cost associated with receiving packets from a free server, but a unit cost is incurred per unit time the costly servers are used. As described above, we can combine all the free servers into one free server from which packets arrive according to a Poisson process of rate R_0. Similarly, we can merge all of the costly servers into one costly server with effective rate of R_c. At any time t, the user has the option to use packets only from the free server or from both the free and the costly servers. In the latter case, the packets arrive according to a Poisson process of rate

$$R_1 = R_0 + R_c.$$

The user's action at time t is denoted by $u_t \in \{0, 1\}$, where $u_t = 0$ if only the free server is used at time t, while $u_t = 1$ if both free and costly servers are used. We assume that the parameters R_0 and R_1 are known at the receiver. Figure 6 illustrates the system model.
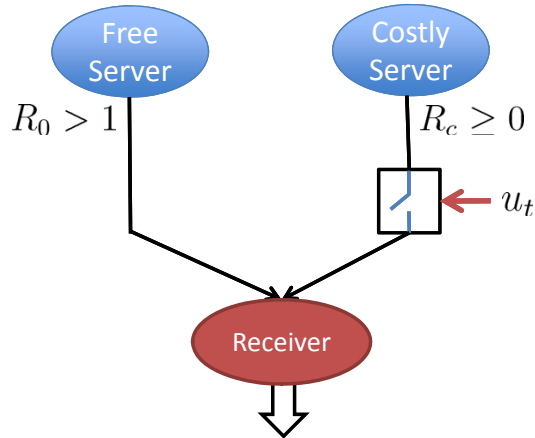

Fig. 6. Streaming frow two classes of servers: costly and free.

The dynamics of the receiver's buffer size (queue-length) Q_t can be described as follows

$$Q_t = D + N_t + \int_0^t u_\tau dN_\tau^c - t, \tag{2}$$

where D is the initial buffer size, N_t Poisson processes of rate R_0 and N_t^c is a Poisson counter of rate R_c which is independent of the process N_t. The last term correspond to the unit rate of media playback.

The user's association (control[2]) policy is formally defined below.

[2]Throughout the rest of this paper, we use the notion of control policy and association policy, interchangeably.

Definition 1. [Control Policy] Let

$$h_t = \{Q_s : 0 \le s \le t\} \cup \{u_s : 0 \le s < t\}$$

denote the history of the buffer sizes and actions up to time t, and $\mathcal{H}$ be the set of all histories for all t. A *deterministic association policy* denoted by π is a mapping $\pi : \mathcal{H} \longmapsto \{0, 1\}$, where at any time t

$$u_t = \pi(h_t) = \begin{cases} 0, & \text{if only the free server is chosen,} \\ 1, & \text{if both servers are chosen.} \end{cases}$$

Denote by Π the set of all such control policies.

We use the initial buffer size D, and interruption probability as QoE metrics. The interruption event occurs when the queue-length Q_t reaches zero. However, this event not only depends on the initial buffer size D, but also on the control policy π. To emphasize this dependency, we denote the interruption probability by

$$p^\pi(D) = \mathbf{Pr}\{\tau_0 < \infty\}, \tag{3}$$

where

$$\tau_0 = \inf\{t : Q_t = 0\}.$$

Definition 2. The policy π is defined to be (D, ϵ)-*feasible* if $p^\pi(D) \le \epsilon$. The set of all such feasible policies is denoted by $\Pi(D, \epsilon)$.

The third metric that we consider in this work is the expected cost of using the costly server which is proportional to the expected usage time of the costly server. For any (D, ϵ), the usage cost of a (D, ϵ)-feasible policy π is given by[3]

$$J^\pi(D, \epsilon) = \mathbf{E}\left[\int_0^{\tau_0} u_t dt \right]. \tag{4}$$

The *value function* or optimal cost function V is defined as

$$V(D, \epsilon) = \min_{\pi \in \Pi(D,\epsilon)} J^\pi(D, \epsilon), \tag{5}$$

and the optimal policy π^* is defined as the optimal solution of the minimization problem in (5).

---

[3]Throughout this work, we use the convention that the cost of an infeasible policy is infinite.

In our model, the user expects to buffer no more than D packets and have an interruption-free experience with probability higher than a desired level $1 - \epsilon$. Note that there are trade-offs among the interruption probability ϵ, the initial buffer size D, and the usage cost. These trade-offs depend on the association policy as well as the system parameters R_0, R_c and F.

Throughout the rest of this work, we study the case that $R_0 > 1$ and the file size F goes to infinity, since the control policies in this case take simpler forms. Moreover, the cost of such control policies in this case provide an upper bound for the finite file size case. The following Lemma summarizes the main result from [5], characterizing the fundamental trade-off between the interruption probability and initial buffering, for a single-server single-receiver system.

Lemma 1. *Consider a single receiver receiving a media stream from a single server according to a Poisson process of rate R. Let D denote the initial buffer size before the playback, and set the playback rate to one. The probability of interruption in media playback is given by*

$$p(D) = e^{-I(R)D}, \tag{6}$$

where $I(R)$ is the largest root of $\gamma(r) = r + R(e^{-r} - 1)$.

We first characterize the region of interest in the space of QoE metrics. In this region, a feasible control policy exists and is non-degenerate. We then use these results to design proper association policies.

Theorem 1. *Let (D, ϵ) be user's QoE requirement when streaming an infinite file from two servers. The arrival rate of the free server is given by $R_0 > 1$, and the total arrival rate when using the costly server is denoted by $R_1 > R_0$. Then*

(a) For any (D, ϵ) such that $D \geq \frac{1}{I(R_0)} \log\left(\frac{1}{\epsilon}\right)$,

$$\min_{\pi \in \Pi} J^\pi(D, \epsilon) = 0.$$

(b) For any (D, ϵ) such that $D < \frac{1}{I(R_1)} \log\left(\frac{1}{\epsilon}\right)$,

$$\min_{\pi \in \Pi} J^\pi(D, \epsilon) = \infty.$$

Proof: Consider the degenerate policy $\pi_0 \equiv 0$. This policy is equivalent to a single-server system with arrival rate $R = R_0$. By Lemma 1, the policy π_0 is (D, ϵ)-feasible for all $D \geq \frac{1}{I(R_0)} \log\left(\frac{1}{\epsilon}\right)$. Note that by (4) this policy does not incur any cost, which results in part (a).

Moreover, for all (D, ϵ) with $D < \frac{1}{I(R_1)} \log\left(\frac{1}{\epsilon}\right)$, there is no (D, ϵ)-feasible policy. This is so since the buffer size under any policy π is stochastically dominated by the one governed by the degenerate policy $\pi_1 \equiv 1$. Hence,

$$p^\pi(D) \geq p^{\pi_1}(D) = \exp(-I(R_1)D) > \epsilon.$$

Using the convention of infinite cost for infeasible policies, we obtain the result in part (b). ∎
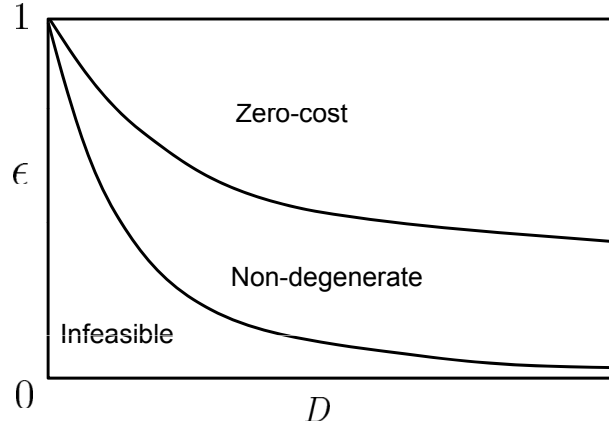

Fig. 7.   Non-degenerate, zero-cost and infeasible regions for QoE metrics (D, ϵ).

For simplicity of notation, let $\alpha_0 = I(R_0)$, and $\alpha_1 = I(R_1)$. By Theorem 1 we focus on the region

$$\mathcal{R} = \left\{ (D, \epsilon) : \frac{1}{\alpha_1} \log\left(\frac{1}{\epsilon}\right) \leq D \leq \frac{1}{\alpha_0} \log\left(\frac{1}{\epsilon}\right) \right\} \tag{7}$$

to analyze the expected cost of various classes of control policies. Figure 7 illustrates a conceptual example of this non-degenerate region as well as the zero-cost and infeasible regions.

IV. DESIGN AND ANALYSIS OF ASSOCIATION POLICIES

In this part, we propose several classes of parameterized control policies. We first characterize the range of the parameters for which the association policy is feasible for a given initial buffer size D and the desired level of interruption probability ϵ. Then, we try to choose the parameters such that the expected cost of the policy is minimized. All of the proofs of the main theorems are included in Appendix A.

A. Off-line Policy

Consider the class of policies where the decisions are made off-line before starting the media streaming. In this case, the arrival process is not observable by the decision maker. Therefore, the user's decision space reduces to the set of deterministic functions $u : \mathbb{R} \to \{0, 1\}$, that maps time into the action space.

Theorem 2. *Let the cost of a control policy be defined as in (4). In order to find a minimum-cost off-line policy, it is sufficient to consider policies of the form:*

$$\pi(h_t) = u_t = \begin{cases} 1, & \text{if } t \leq t_s \\ 0, & \text{if } t > t_s, \end{cases} \tag{8}$$

which parameterized are by a single parameter $t_s \geq 0$.

Proof: In general, any off-line policy π consists of multiple intervals in which the costly server is used. Consider an alternative policy π' of the form of (8) where $t_s = J^\pi$. By definition of the cost function in (4), the two policies incur the same cost. Moreover, the buffer size process under policy π is stochastically dominated by the one under policy π', because the policy π' counts the arrivals from the costly server earlier, and the arrival process is stationary. Hence, the interruption probability of π' is not larger than that of π. Therefore, for any off-line policy, there exists another off-line policy of the form given by (8). ∎

Theorem 3. *Consider the class of off-lines policies of the form (8). For any $(D, \epsilon) \in \mathcal{R}$, the policy π defined in (8) is feasible if*

$$t_s \geq t_s^* = \frac{R_0}{R_1 - R_0} \left[\frac{1}{\alpha_0} \log \left(\frac{1}{\epsilon - e^{-\alpha_1 D}} \right) - D \right]. \tag{9}$$

Note that obtaining the optimal off-line policy is equivalent to finding the smallest t_s for which the policy is still feasible. Therefore, t_s^* given in (9) provides an upper bound on the minimum cost of an off-line policy. Observe that t_s^* is almost linear in D for all (D, ϵ) that is not too close to the lower boundary of region $\mathcal{R}$. As (D, ϵ) gets closer to the boundary, t_s^* and the expected cost grows to infinity, which is in agreement with Theorem 1. In this work, we pick t_s^* as a benchmark for comparison to other policies that we present next.

B. Online Safe Policy

Let us now consider the class of online policies where the decision maker can observe the buffer size history. Inspired by the structure of the optimal off-line policies, we first focus on a *safe* control policy in which, in order to avoid interruptions, the costly server is used at the beginning until the buffer size reaches a certain threshold, after which the costly server is never used. This policy is formally defined below.

Definition 3. The online safe policy π^S parameterized by the threshold value S is given by

$$\pi^S(h_t) = \begin{cases} 1, & \text{if } t \leq \tau_s \\ 0, & \text{if } t > \tau_s, \end{cases} \tag{10}$$

where $\tau_S = \inf\{t \geq 0 : Q_t \geq S\}$.

Theorem 4. *Let π^S be the safe policy defined in Definition 3. For any $(D, \epsilon) \in \mathcal{R}$, the safe policy is feasible if*

$$S \geq S^* = \frac{1}{\alpha_0} \log\left(\frac{1}{\epsilon - e^{-\alpha_1 D}}\right). \tag{11}$$

Moreover,

$$\min_{S \geq S^*} J^{\pi^S}(D, \epsilon) = J^{\pi^{S^*}}(D, \epsilon) = \frac{1}{R_1 - 1}\left[\frac{1}{\alpha_0} \log\left(\frac{1}{\epsilon - e^{-\alpha_1 D}}\right) - D + \xi\right],$$

where $\xi \in [0, 1)$.

Let us now compare the online safe policy π^{S^*} with the off-line policy defined in (8) with parameter t_s^* as in (9). We observe that the cost of the online safe policy is almost proportional to that of the off-line policy, where the cost ratio of the off-line policy to that of the online safe policy is given by

$$\frac{R_0(R_1 - 1)}{R_1 - R_0} = 1 + \frac{R_1(R_0 - 1)}{R_1 - R_0} > 1.$$

Note that the structure of both policies is the same, i.e, both policies use the costly server for a certain period of time and then switch back to the free server. As suggested here, the advantage of observing the buffer size allows the online policies to avoid excessive use of the costly server when there are sufficiently large number of arrivals from the free server. In the following, we present another class of online policies.

C. Online Risky Policy

In this part, we study a class of online policies where the costly server is used only if the buffer size is below a certain threshold. We call such policies "risky" as the risk of interruption is spread out across the whole trajectory, unlike the "safe" policies. Further, we constrain risky policies to possess the property that the action at a particular time should only depend on the buffer size at that time, i.e., such policies are *stationary Markov* with respect to buffer size as the state of the system. The risky policy is formally defined below.

Definition 4. The online risky policy π^T parameterized by the threshold value T is given by

$$\pi^T(h_t) = \pi^T(Q_t) = \begin{cases} 1, & \text{if } 0 < Q_t < T \\ 0, & \text{otherwise.} \end{cases} \tag{12}$$

Theorem 5. *Let π^T be the risky policy defined in Definition 4. For any $(D, \epsilon) \in \mathcal{R}$, the policy π^T is feasible if the threshold T satisfies*

$$T \geq T^* = \begin{cases} \frac{1}{\alpha_1 - \alpha_0}\left[\log\left(\frac{\beta}{\epsilon}\right) - \alpha_0 D\right], & \text{if } D \geq \bar{D}, \\ \frac{1}{\alpha_1}\log\left(\frac{\epsilon + \beta(1 - e^{-\alpha_1 D}) - 1}{\epsilon - e^{-\alpha_1 D}}\right), & \text{if } D \leq \bar{D}, \end{cases} \tag{13}$$

where $\beta = \frac{\alpha_1}{\alpha_0(1 - \frac{\alpha_0}{2})}$ and $\bar{D} = \frac{1}{\alpha_1}\log\left(\frac{\beta}{\epsilon}\right)$.

Theorem 5 facilitates the design of risky policies with a single-threshold structure, for any desired initial buffer size D and interruption probability ϵ. For a fixed ϵ, when D increases, T^* (the design given by Theorem 5) decreases to zero. On the other hand, if D decreases to $\frac{1}{\alpha_1}\log\left(\frac{1}{\epsilon}\right)$ (the boundary of $\mathcal{R}$), the threshold T^* quickly increases to infinity, i.e., the policy does not switch back to the free server unless a sufficiently large number of packets is buffered. Figure 8 plots T^* and D as a function of D for a fixed ϵ. Observe that, for large range of D, $T^* \leq D$, i.e., the costly server is not initially used. In this range, owing to the positive drift of Q_t, the probability of ever using the costly server exponentially decreases in $(D - T^*)$.

Next, we compute bounds on the expected cost of the online risky policy and compare with the previously proposed policies.

Theorem 6. *For any $(D, \epsilon) \in \mathcal{R}$, consider an online risky policy π^{T^*} defined in Definition 4, where the threshold T^* is given by (13) as function of D and ϵ. If $D \geq \bar{D}$ then*

$$J^{\pi^{T^*}}(D, \epsilon) \leq \frac{\beta}{\alpha_1(R_1 - 1)}e^{-a_0(D - T^*)}, \tag{14}$$
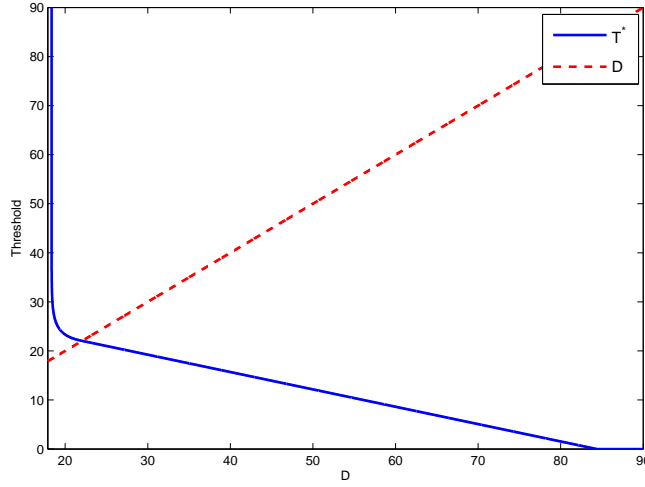
Fig. 8. The switching threshold of the online risky policy as a function of the initial buffer size for $\epsilon = 10^{-3}$ (See Theorem 5).

and if $D \leq \bar{D}$

$$J^{\pi^{T^*}}(D, \epsilon) \leq \frac{1 - e^{-\alpha_1 D}}{(R_1 - 1)(1 - e^{-\alpha_1 T^*})}\left(T^* + 1 + \frac{\beta}{\alpha_1}\right) - \frac{D}{R_1 - 1}, \qquad (15)$$

where $\beta = \frac{\alpha_1}{\alpha_0(1 - \frac{\alpha_0}{2})}$ and $\bar{D} = \frac{1}{\alpha_1}\log\left(\frac{\beta}{\epsilon}\right)$.

In the following, we compare the expected cost of the presented policies using numerical methods, and illustrate that the bounds derived in Theorems 3, 4 and 6 on the expected cost function are close to the exact value.

D. Performance Comparison

Figure 9 compares the expected cost functions of the off-line, online safe and online risky policies as a function of the initial buffer size D, when the interruption probability is fixed to $\epsilon = 10^{-3}$, the arrival rate from the free server is $R_0 = 1.05$, and the arrival rate from the costly server is $R_c = R_1 - R_0 = 0.15$. We plot the bounds on the expected cost given by Theorems 3, 4 and 6 as well as the expected cost function numerically computed by the Monte-Carlo method. Figure 9 shows that the analytical bounds we computed for the expected cost of various control policies closely match the exact cost functions computed via simulations.
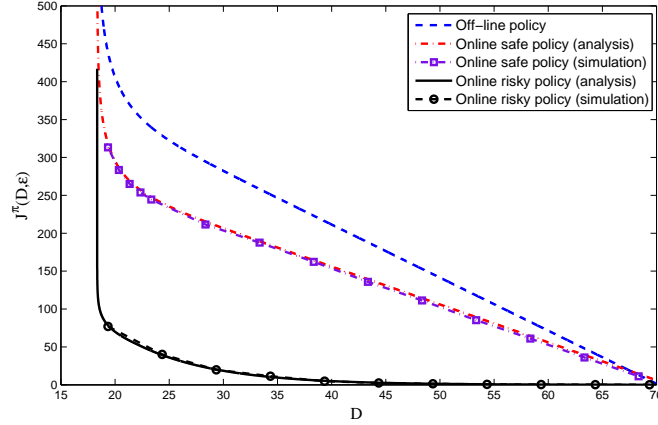
Fig. 9. Expected cost (units of time) of the presented control policies as a function of the initial buffer size for interruption probability $\epsilon = 10^{-3}$. The analytical bounds are given by Theorems 3, 4 and 6.

Observe that the expected cost of the risky policy is significantly smaller that both online safe and off-line policies. For example, the risky policy allows us to decrease the initial buffer size from 70 to 20 with an average of $70 \times 0.15 \approx 10$ extra packets from the costly server. The expected cost in terms of the number packets received from the costly server is 43 and 61 for the online safe and off-line policy, respectively.

Moreover, note that it is merely the existence of the costly server as a backup that allows us to improve the user's quality of experience without actually using too many packets from the costly server. For example, observe that the risky policy satisfies QoE metrics of $D = 35$ and $\epsilon = 10^{-3}$, by only using on average about one extra packet from the costly server. However, without the costly server, in order to decrease the initial buffer size from 70 to 35, the interruption probability has to increase from 10^{-3} to about 0.03 (see Lemma 1).

V. DYNAMIC PROGRAMMING APPROACH

In this section, we present a characterization of the optimal association policy in terms of the Hamilton-Jacobi-Bellman (HJB) equation. Note that because of the probabilistic constraint over the space of sample paths of the buffer size, the optimal policy is not necessarily Markov with respect to the buffer size as the state of the system. We take a similar approach as in [6] where by expanding the state space, a Bellman equation is provided as the optimality condition of an

MDP with probabilistic constraint. In particular, consider the pair (Q_t, p_t) as the state variable, where Q_t denotes the buffer size and p_t represents the desired level of interruption probability given the information at time t. Note that p_t is a Martingale by definition [28]. The evolution of Q_t is governed by the following stochastic differential equation

$$dQ_t = -dt + dN^u, \quad Q_0 = D, \tag{16}$$

where N^u is a Poisson counter with rate $R_{u_t} = R_0 + u_t \cdot R_c$. For any $(D, \epsilon) \in \mathcal{R}$ and any optimal policy π, the constraint $p^\pi(D) \leq \epsilon$ is active. Hence, we consider the sample paths of p_t such that $p_0 = \epsilon$. Moreover, we have $\mathbf{E}[p_t] = \epsilon$ for all t, where the expectation is with respect to the Poisson jumps. Let $dp_t = \hat{p}_t - p_t$ be the change in state p, if a Poisson jump occurs in an infinitesimal interval of length dt. Also, let $dp_t = dp_0$ be the change in state p if no jump occurs. Therefore,

$$0 = \mathbf{E}[dp_t] = R_{u_t} dt(\hat{p}_t - p_t) + (1 - R_{u_t} dt) dp_0.$$

By solving the above equation for dp_0, we obtain the evolution of p_t as a function of the control process $\hat{p}_t$ and u_t:

$$dp_t = (p_t - \hat{p}_t)(R_{u_t} dt - dN^u), \quad p_0 = \epsilon. \tag{17}$$

Similarly to the arguments of Theorem 2 of [6], by principle of optimality we can write the following dynamic programming equation

$$V(Q, p) = \min_{u \in \{0,1\}, \hat{p} \in [0,1]} \left\{ u dt + \mathbf{E}[V(Q + dQ, p + dp)] \right\}. \tag{18}$$

If V is continuously differentiable, by Itō's Lemma for jump processes, we have

$$V(Q + dQ, p + dp) - V(Q, p) = \frac{\partial V}{\partial Q}(-dt) + \frac{\partial V}{\partial p} \cdot (p - \hat{p}) R_u dt + \left(V(Q + 1, \hat{p}) - V(Q, p) \right) dN^u,$$

which implies the following HJB equation after dividing (18) by dt and taking the limit as t goes to zero:

$$\frac{\partial V(Q, p)}{\partial Q} = \min_{u \in \{0,1\}, \hat{p} \in [0,1]} \left\{ u + \frac{\partial V}{\partial p} \cdot (p - \hat{p}) R_u + R_u \left(V(Q + 1, \hat{p}) - V(Q, p) \right) \right\}$$

The optimal policy π is obtained by characterizing the optimal solution of the partial differential equation in (19) together with the boundary condition $V(Q, 1) = 0$. Since such equations are in general difficult to solve analytically, we use the *guess and check* approach, where we

propose a candidate for the value function and verify that it nearly satisfies the HJB equation almost everywhere. For any $(Q, p) \in \mathcal{R}$, define

$$T(Q,p) = \begin{cases} \frac{1}{\alpha_1 - \alpha_0}\left[\log\left(\frac{\theta}{p}\right) - \alpha_0 Q\right], & \text{if } Q \geq \frac{1}{a_1}\log\left(\frac{\theta}{p}\right), \\ \frac{1}{\alpha_1}\log\left(\frac{p + \theta(1 - e^{-\alpha_1 Q}) - 1}{p - e^{-\alpha_1 Q}}\right), & \text{otherwise,} \end{cases} \tag{19}$$

where $\theta = \frac{\alpha_1}{\alpha_0}$. The candidate solution for HJB equation (19) is given by

$$\bar{V}(Q,p) = \frac{1}{\alpha_0(1 - \frac{\alpha_0}{2})(R_1 - 1)} e^{-a_0(Q - T(Q,p))}, \tag{20}$$

when $Q \geq \frac{1}{a_1}\log\left(\frac{\theta}{p}\right)$, and

$$\bar{V}(Q,p) = \frac{p + \theta(1 - e^{-\alpha_1 Q}) - 1}{(R_1 - 1)(\theta - 1)}\left(T(Q,p) + \frac{\beta}{\alpha_1}\right) - \frac{Q}{R_1 - 1}, \tag{21}$$

when $Q < \frac{1}{a_1}\log\left(\frac{\theta}{p}\right)$. Note that the candidate solution is derived from the structure of the expected cost of the risky policy (cf. Theorem 6). We may verify that $\bar{V}$ satisfies the HJB equation (19) for all (Q, p) such that $Q \geq \frac{1}{a_1}\log\left(\frac{\theta}{p}\right)$ or $Q \leq \frac{1}{a_1}\log\left(\frac{\theta}{p}\right) - 1$, but for other (Q, p) the HJB equation is only approximately satisfied. This is due to the discontinuity of the queue-length process which does not allow us to exactly match the expected cost starting from below the threshold with the one starting from above the threshold. Therefore, owing to approximate characterization of the cost of the risky policy, we may not prove or disprove optimality of this policy.

In the following, we use a fluid model to provide an *exact* characterization of the optimal control policy using appropriate HJB equation. We show that the optimal policy takes a threshold structure similarly to the online risky policy.

VI. OPTIMAL ASSOCIATION POLICY FOR A FLUID MODEL

Thus far, we concentrated on design and analysis of various network association policies in an uncertain environment, where network uncertainties are modeled using a Poisson arrival process. We provided closed-form approximations of the cost of different policies. However, an exact analytical solution is required to prove optimality of the risky policy. This is particularly challenging, since an exact distribution of threshold over-shoots is desired, owing to the dis-continuous nature of the Poisson process. In this part, we exploit a second-order approximation of the Poisson process [29] and model the receiver's buffer size using a controlled Brownian motion with drift.

Consider the system model as in Figure 6, with following queue-length dynamics at the receiver:

$$dQ_t = (R_{u_t} - 1)dt + dW_t, \quad Q_0 = D, \tag{22}$$

where W_t is the Wiener process, $u_t \in \{0, 1\}$ is the receiver's decision at time t on using the free or costly server. As in the preceding part, we assume that the media file size F is infinite and $R_1 > R_0 > 1$.

Define the control policy (network association policy) as in Definition 1. The goal is to find a feasible policy that minimizes the usage cost defined in (4) such that the interruption probability $p^\pi(D)$ defined in (3) is at most ϵ. As in the previous part, the set of feasible policies and the value function is given by Definition 2 and (5), respectively. The following lemma is the counterpart of Lemma 1 for the fluid model.

Lemma 2. *Let $\pi_i \equiv i$ denote a degenerate policy, for $i \in \{0, 1\}$. The interruption probability for such policy is given by*

$$p^{\pi_i}(D) = e^{-\theta_i D}, \quad \text{for all } D \geq 0, \tag{23}$$

where $\theta_i = 2(R_i - 1)$, for $i \in \{0, 1\}$.

Proof: See Appendix B. ∎

First, we provide a characterization of the optimal policy via a Hamilton-Jacobi-Bellman equation. As in Section V, we expand the state variables to (Q, p), where Q is the queue-length with dynamics given by 22, and p is the desired interruption probability. Using Martingale representation theorem [28], we may write the dynamics of p as follows:

$$dp_t = \hat{p}_t \, dW_t, \quad p_0 = \epsilon, \tag{24}$$

where $\hat{p}_t$ is a predictable process which is adapted with respect to natural filtration of the history process. In this work, we focus on the control processes that are Markovian with respect to the state process (Q_t, p_t). Therefore, using the principal of optimality, we may write the following dynamic programming equation:

$$V(Q, p) = \min_{(u, \hat{p}) \in \{0, 1\} \times \mathbb{R}} \{ u dt + \mathbf{E}[V(Q + dQ, p + dp)] \}, \tag{25}$$

where $(u, \hat{p})$ are the control actions. For a twice differentiable function V, we may exploit Itō's Lemma to get

$$
\begin{aligned}
dV(Q, p) \;=\;& \frac{\partial V}{\partial Q}dQ + \frac{\partial V}{\partial p}dp + \frac{1}{2}\frac{\partial^2 V}{\partial Q^2}(dQ)^2 + \frac{1}{2}\frac{\partial^2 V}{\partial p^2}(dp)^2 + \frac{\partial^2 V}{\partial Q \partial p}dQdp \\
\overset{(a)}{=}\;& \frac{\partial V}{\partial Q}((R_u - 1)dt + dW) + \frac{\partial V}{\partial p}(\hat{p}\, dW) \\
&+ \frac{1}{2}\frac{\partial^2 V}{\partial Q^2}dt + \frac{1}{2}\frac{\partial^2 V}{\partial p^2}(\hat{p})^2 dt + \frac{\partial^2 V}{\partial Q \partial p}\hat{p}dt,
\end{aligned}
$$

where (a) follows from state dynamics in (22) and (24). Replacing the above equation back in (25), and taking the expectation with respect to dW, and limit as dt tends to zero, we obtain the following HJB equation

$$
0 = \min_{(u, \hat{p}) \in \{0,1\} \times \mathbb{R}} \left\{ u + \frac{\partial V}{\partial Q}(R_u - 1) + \frac{1}{2}\frac{\partial^2 V}{\partial Q^2} + \frac{1}{2}\frac{\partial^2 V}{\partial p^2}(\hat{p})^2 + \frac{\partial^2 V}{\partial Q \partial p}\hat{p} \right\}. \tag{26}
$$

Note that we require the following boundary conditions for the value function:

$$
V(Q, 1) = V(0, p) = 0, \quad \text{for all } Q \geq 0, 0 \leq p \leq 1 \tag{27}
$$

Providing an analytical solution for the partial differential equation in (26) is often challenging. However, we may take a guess and check approach and use a threshold policy as the basis of our guess. Note that we need to verify the HJB equation for the set of state variables that reachable by a feasible policy. In particular, in light of Lemma 2, it is clear that for $p \leq e^{-\theta_1 Q}$, there is no feasible policy and the value function $V(Q, p) = \infty$. Moreover, for all $p \geq e^{-\theta_0 Q}$, observe that the degenerate policy $\pi_0 \equiv 0$ is optimal, $V(Q, p) = 0$ which also satisfies the HJB equation and the boundary conditions. Therefore, we focus on the non-degenerate region

$$
\mathcal{R} = \left\{ (Q, p) : Q \geq 0, e^{-\theta_1 Q} < p < e^{-\theta_0 Q} \right\}. \tag{28}
$$

Figure 7 illustrates a conceptual example of this non-degenerate region. In the following, we first define a threshold policy similar to the risky policy of Definition 4, and present a closed-form characterization of its cost function. Then, we show that, for a proper choice of the threshold the associated cost function satisfies the HJB equation in (26), and the optimal solution of the minimization problem in (26) coincides with the threshold policy. Hence, we establish the optimality of the proposed policy.

Theorem 7. *Let π^T be the threshold policy as in Definition 4, parameterized with threshold value T. Also, let the queue-length dynamics be governed by (22). Then, the interruption probability for this policy is given by*

$$p^T(D) = \begin{cases} e^{-\theta_1 D} + p(T)(1 - \frac{\theta_0}{\theta_1})(1 - e^{-\theta_1 D}), & 0 \leq D \leq T \\ p(T)e^{-\theta_0(D-T)}, & D \geq T, \end{cases} \tag{29}$$

where

$$p(T) = \frac{e^{-\theta_1 T}}{\frac{\theta_0}{\theta_1} + (1 - \frac{\theta_0}{\theta_1})e^{-\theta_1 T}},$$

and $\theta_i \triangleq 2(R_i - 1)$, for $i \in \{0, 1\}$.

Proof: See Appendix B. ■

Corollary 1. *Let π^T be the threshold policy as in Definition 4. Then, the policy π^T is (D, ϵ)-feasible (cf. Definition 2) for the following choices of the threshold T:*

1) *For all $\epsilon \geq e^{-\theta_0 D}$, let $T = 0$.*
2) *For all $e^{-\theta_1 D} \leq \epsilon \leq e^{-\theta_0 D}$, let $T = T(D, \epsilon)$ be the unique solution of $p^T(D) = \epsilon$, where $p^T(D)$ is given by (29).*
3) *For all other ϵ, there exists no such T.*

Proof: The proof directly follows from the characterization of the interruption probability in Theorem 7, noting the fact that $p^T(D) \in [e^{-\theta_1 D}, e^{-\theta_0 D}]$ is monotonically decreasing in T. ■

Next, we provide a exact characterization of the expected cost of the threshold policy π^T for a given threshold T. This allows us to obtain a proper candidate solution for the HJB equation.

Theorem 8. *Let π^T be the threshold policy as in Definition 4. Define $J^T(D)$ as the expected cost associated with policy π^T given the initial condition D for queue-length dynamics (22) and threshold T. The* cost-to-go *function $J^T(D)$ is given by*

$$J^T(D) = \begin{cases} e^{-\theta_0(D-T)}J(T), & D \geq T \\ (J(T) + \frac{2}{\theta_1}T)\frac{1-e^{-\theta_1 D}}{1-e^{-\theta_1 T}} - \frac{2}{\theta_1}D, & D \leq T, \end{cases} \tag{30}$$

where

$$J(T) = \frac{\frac{2}{\theta_1^2}\left[1 - (1 + \theta_1 T)e^{-\theta_1 T}\right]}{\frac{\theta_0}{\theta_1} + (1 - \frac{\theta_0}{\theta_1})e^{-\theta_1 T}}. \tag{31}$$

Proof: See Appendix B. ■

The following theorem provide a candidate for value function and verify the optimality condition given by HJB equation in (26).

Theorem 9. *For all $(Q, p) \in \mathcal{R}$, define*

$$V(Q, p) = J^{T(Q,p)}(Q), \tag{32}$$

where $J^T(\cdot)$ is defined in (30), $\mathcal{R}$ is defined in (28), and $T(Q, p)$ is the unique solution of

$$p^T(Q) = p.$$

Then, the HJB equation (26) and boundary condition (27) hold for all $(Q, p) \in \mathcal{R}$.

Proof: See Appendix B. ∎

Theorem 9 verifies that the value function $V(Q, p)$ given by (32) is indeed the optimal cost function defined in (5). Furthermore, we can conclude that the policy $\pi^*(Q, p)$ achieving the minimum in the HJB equation (26) is optimal. In general, the optimal policy depends on both state variables (Q, p) and is not Markov with respect to Q. In the following, we show that the state trajectory steered by the optimal policy is limited to a one-dimensional manifold and the threshold policy π^T is optimal for all $(Q, p) \in \mathcal{R}$. Recall the policy π^T boils down to the optimal policy $\pi_0 \equiv 0$ for all other admissible states, by using threshold value $T = 0$.

Theorem 10. *Let $\pi^*(Q, p)$ attain the minimum in the HJB equation (26) for any $(Q, p) \in \mathcal{R}$. Let (Q_t^*, p_t^*) denote the state trajectory given the initial condition (D, ϵ), under the control trajectory $(u_t^*, \hat{p}_t^*) = \pi^*(Q_t^*, p_t^*)$. Then, the state trajectory is limited to a one-dimensional invariant manifold $\mathcal{M}(D, \epsilon)$, where*

$$\mathcal{M}(D, \epsilon) = \big\{ (Q, p) : p = p^{T(D,\epsilon)}(Q) \big\}, \tag{33}$$

where $T(D, \epsilon)$ is the solution of $p^T(D) = \epsilon$, and $p^T(\cdot)$ is defined in (29). Moreover, the optimal policy $\pi^*(Q, p)$ coincides with the threshold policy $\pi^{T(D,\epsilon)}(Q)$.

Proof: See Appendix B. ∎

Figure 10 illustrates a conceptual figure describing the intuition behind Theorem 10. Observe that the optimal policy satisfying the HJB equation, divides the feasible state space into two sub-regions corresponding to $u^* = 0$ and $u^* = 1$, i.e., the policy switches costly server on/off when the state of the system crosses the boundary between these sub-regions. Theorem 10 states

that for any initial condition $(Q_0, p_0) = (D, \epsilon)$ the state trajectory lies on a one-dimensional manifold $\mathcal{M}(D, \epsilon)$. Figure 10 illustrates these manifolds for different initial conditions. Since the state trajectory is limited to a one-dimensional space, the decision of switching to the costly server merely depends on the queue-length process. The proof of Theorem 10 in Appendix B, shows that the queue-length at the switch point for each manifold $\mathcal{M}(D, \epsilon)$ coincides with the threshold $T(D, \epsilon)$ specified in Corollary 1.
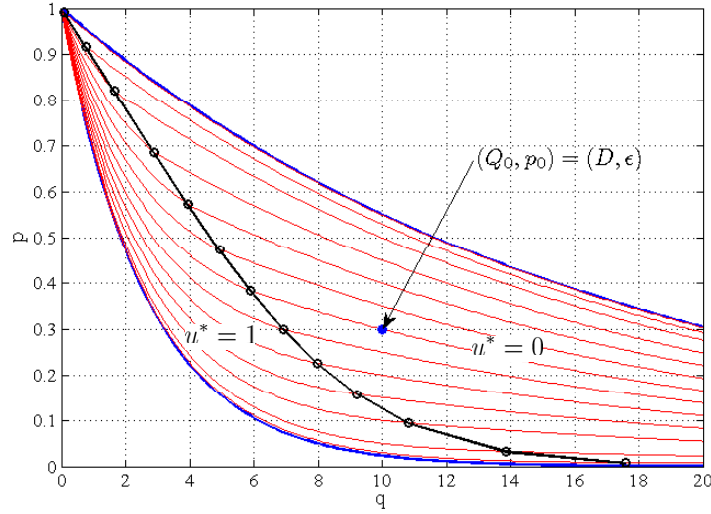

Fig. 10. Trajectory of the optimal policy lies on a one-dimensional manifold.

VII. CONCLUSIONS AND FUTURE WORK

We presented a new framework for studying media streaming systems in volatile environments, with focus on quality of user experience. We proposed two intuitive metrics that essentially capture the notion of *delay* from the end-user's point of view. The proposed metrics in the context of media streaming, are initial buffering delay, and probability of interruption in media playback. These metrics are tractable enough to be used as a benchmark for system design.

We first addressed the problem of streaming in a technology-heterogeneous multi-server system. The main challenge in multi-server systems is inefficiencies in multi-path streaming due to duplicate packet reception. This issue can also significantly complicates the analysis. We proposed random linear network coding as the solution to this challenge. By sending random linear combination of packets, we remove the notion of identity from packets and hence,

guarantee that no packet is redundant. Using this approach allows us to significantly simplify the flow control of multi-path streaming scenarios, and model heterogeneous multi-server systems as a single-server system.

Equipped with tools provided by network coding, we added another level of complexity to the multi-server system. We used our framework to study multi-server systems when the access cost varies across different servers. Our objective was to investigate the trade-offs between the network usage cost and the user's QoE requirements parameterized by initial waiting time and allowable probability of interruption in media playback. For a Poisson arrival model, we analytically characterized and compared the expected cost of both off-line and online policies, finally showing that a threshold-based online risky policy achieves the lowest cost. The threshold policy uses the costly server if and only if the receiver's buffer is below a certain threshold. Moreover, we observed that even rare but properly timed usage of alternative access technologies significantly improves user experience without any bandwidth over-provisioning.

We formulated the access cost minimization problem as a Markov decision problem with probabilistic constraints, and characterized the optimal policy by the HJB equation. For a fluid approximation model, we established the optimality of a threshold-based online policy in the class of deterministic Markov policies using the HJB equation as a verification method.

The framework that we have developed in this work can also be used to design adaptive resolution streaming systems that not only depend on the channel conditions but also on the delay requirement of the application, which is captured by the queue-length at the receiver. As for other extensions of this work, we would like to study more accurate models of channel variations such as the two-state Markov model due to Gillbert and Elliot. In this work we focused on deterministic network association policies. Another extension of this work would consist of studying randomized control policies.

APPENDIX A
ANALYSIS OF THE CONTROL POLICIES FOR THE POISSON ARRIVAL MODEL

Proof of Theorem 3. By Definition 2, we need to show that $p^\pi(D) \leq \epsilon$. By a union bound on the interruption probability, it is sufficient to verify

$$\mathbf{Pr}\Big(\min_{0 \leq t \leq t_s} Q_t \leq 0 \big| Q_0 = D\Big) + \mathbf{Pr}\Big(\min_{t > t_s} Q_t \leq 0 \big| Q_0 = D\Big) \leq \epsilon. \tag{34}$$

In the interval $[0, t_s]$, Q_t behaves as in a single-server system with rate R_1. Hence, by Lemma 1 we get

$$\mathbf{Pr}\left(\min_{0 \leq t \leq t_s} Q_t \leq 0 \big| Q_0 = D\right) \leq e^{-\alpha_1 D}. \tag{35}$$

For the second term in (34), we have

$$
\begin{aligned}
\mathbf{Pr}\left(\min_{t > t_s} Q_t \leq 0 \big| Q_0 = D\right) &= \sum_{q=D-t_s}^{\infty} \mathbf{Pr}\left(\min_{t > t_s} Q_t \leq 0 \big| Q_{t_s} = q\right) \mathbf{Pr}(Q_{t_s} = q) \\
&\overset{(a)}{\leq} \sum_{q=D-t_s}^{\infty} e^{-\alpha_0 q} \mathbf{Pr}(Q_{t_s} = q) \\
&= \sum_{k=0}^{\infty} e^{-\alpha_0(D+k-t_s)} \mathbf{Pr}(N_{t_s} + N_{t_s}^c = k) \\
&\overset{(b)}{=} \sum_{k=0}^{\infty} e^{-\alpha_0(D+k-t_s)} \frac{e^{-R_1 t_s}(R_1 t_s)^k}{k!} \\
&= e^{-\alpha_0(D-t_s)+R_1 t_s(e^{-\alpha_0}-1)} \sum_{k=0}^{\infty} \frac{e^{-R_1 t_s e^{-\alpha_0}}(R_1 t_s e^{-\alpha_0})^k}{k!} \\
&= \exp\left(-\alpha_0(D-t_s) + R_1 t_s(e^{-\alpha_0} - 1)\right) \cdot 1 \\
&\overset{(c)}{=} \exp\left(-\alpha_0(D-t_s) + R_1 t_s(-\frac{\alpha_0}{R_0})\right) \\
&\overset{(d)}{\leq} \epsilon - e^{-\alpha_1 D},
\end{aligned}
$$

where (a) follows from Lemma 1 and the fact that $u_t = 0$, for $t \geq t_s$. (b) is true because $N_{t_s} + N_{t_s}^c$ is a Poisson random variable with mean $R_1 t_s$. (c) holds since $\alpha_0 = I(R_0)$ is the root of $\gamma(r) = r + R_0(e^{-r} - 1)$. Finally, (d) follows from the hypothesis of the theorem.

By combining the above bounds, we may verify (34) which in turns proves feasibility of the proposed control policy. ∎

Proof of Theorem 4. Similarly to the proof of Theorem 3, we need to show that the total probability of interruption before and after crossing the threshold S is bounded from above by ϵ. Observe that for any realization of τ_S the bound in (35) still holds. Further, since the costly server is not used after crossing the threshold and $Q_{\tau_S} \geq S$, Lemma 1 implies

$$\mathbf{Pr}\left(\min_{t > \tau_S} Q_t \leq 0 \big| Q_0 = D\right) \leq e^{-\alpha_0 S} \leq \epsilon - e^{-\alpha_1 D}, \tag{36}$$

where the second inequality follows from (11). Finally, combining (35) and (36) gives $p^{\pi^S}(D) \leq \epsilon$, which is the desired feasibility result.

For the second part, first observe that $J^{\pi^S}(D, \epsilon) = \mathbf{E}[\tau_S]$. In order to cross a threshold $S \geq S^*$, the threshold S^* must be crossed earlier, because $Q_0 = D \leq S^*$. Hence, τ_S stochastically dominates τ_S^*, implying

$$J^{\pi^S}(D, \epsilon) = \mathbf{E}[\tau_S] \geq \mathbf{E}[\tau_{S^*}] = J^{\pi^{S^*}}(D, \epsilon), \quad \text{for all } S \geq S^*.$$

It only remains to compute $\mathbf{E}[\tau_{S^*}]$. It follows from Wald's identity or Doob's optional stopping theorem [28] that

$$D + (R_1 - 1)\mathbf{E}[\tau_{S^*}] = \mathbf{E}[Q_{\tau_{S^*}}] = S^* + \xi, \tag{37}$$

where $\xi \in [0, 1)$ because the jumps of a Poisson process are of units size, and hence the overshoot size when crossing a threshold is bounded by one, i.e., $S^* \leq Q_{\tau_{S^*}} < S^* + 1$. Rearranging the terms in (37) and plugging the value of S^* from (11) immediately gives the result. ∎

Lemma 3. *Let Q_t be the buffer size of a single-server system with arrival rate $R > 1$. Let the initial buffer size be D and for any $T \geq D > 0$ define the following stopping times*

$$\tau_T = \inf\{t > 0 : Q_t \geq T\}, \quad \tau_e = \inf\{t \geq 0 : Q_t \leq 0\}. \tag{38}$$

Then

$$\mathbf{Pr}(\tau_e > \tau_T) = \frac{1 - e^{-I(R)D}}{1 - \mathbf{E}[e^{-I(R)Q_{\tau_T}}|\tau_e > \tau_T]}, \tag{39}$$

where $I(R)$ is defined in Lemma 1.

Proof: Let $Y(t) = e^{-I(R)Q_t}$. We may verify that $Y(t)$ is a Martingale and uniformly integrable. Also, define the stopping time $\tau = \min\{\tau_T, \tau_e\}$. Since $R > 1$, we have $\mathbf{Pr}(\tau \geq t) \leq \mathbf{Pr}(0 < Q_t < T) \to 0$, as $t \to \infty$. Hence, $\tau < \infty$ almost surely. Therefore, we can employ Doob's optional stopping theorem [28] to write

$$
\begin{aligned}
e^{-I(R)D} &= \mathbf{E}[Y(0)] = \mathbf{E}[Y(\tau)] \\
&= \mathbf{Pr}(\tau_e \leq \tau_T) \cdot 1 \\
&\quad + \mathbf{Pr}(\tau_e > \tau_T)\mathbf{E}[e^{-I(R)Q_{\tau_T}}|\tau_e > \tau_T].
\end{aligned}
$$

The claim immediately follows from the above relation after rearranging the terms.

∎

Proof of Theorem 5. Let us first characterize the interruption probability of the policy π^T when the initial buffer size is $D = T$. In this case, by definition of π^T the behavior of Q_t is initially the same as a single-server system with rate R_1 until the threshold T is crossed. Hence,

$$
\begin{aligned}
p^{\pi^T}(T) &= \mathbf{Pr}\Big(\min_{t \geq 0} Q_t \leq 0 \big| Q_0 = T\Big) \\
&= \mathbf{Pr}(\tau_e < \tau_T) \cdot 1 \\
&\quad + \mathbf{Pr}(\tau_T < \tau_e)\mathbf{Pr}\Big(\min_{t \geq \tau_T} Q_t \leq 0 \big| \tau_T < \tau_e, Q_0 = T\Big) \\
&= \frac{e^{-\alpha_1 T} - \mathbf{E}[e^{-\alpha_1 Q_{\tau_T}} | \tau_e > \tau_T]}{1 - \mathbf{E}[e^{-\alpha_1 Q_{\tau_T}} | \tau_e > \tau_T]} \\
&\quad + \frac{\big(1 - e^{-\alpha_1 T}\big)\mathbf{Pr}\Big(\min_{t \geq \tau_T} Q_t \leq 0 \big| \tau_T < \tau_e, Q_0 = T\Big)}{1 - \mathbf{E}[e^{-\alpha_1 Q_{\tau_T}} | \tau_e > \tau_T]},
\end{aligned}
\tag{40}
$$

where the last equality follows directly from Lemma 3. Further, we have

$$
\begin{aligned}
\mathbf{Pr}\Big(\min_{t \geq \tau_T} Q_t \leq 0 \big| \tau_T < \tau_e, Q_0 = T\Big) &= \int_T^{T+1} \mathbf{Pr}\Big(\min_{t \geq \tau_T} Q_t \leq 0 \big| Q_{\tau_T}\Big) d\mu(Q_{\tau_T}) \\
&\overset{(a)}{=} \int_T^{T+1} \mathbf{Pr}\Big(\min_{t \geq 0} Q_t \leq 0 \big| Q_0\Big) d\mu(Q_0) \\
&\overset{(b)}{=} \int_T^{T+1} \mathbf{Pr}\Big(\min_{t \geq 0} Q_t \leq 0 \big| \min_{t \geq 0} Q_t \leq T, Q_0\Big) \mathbf{Pr}\Big(\min_{t \geq 0} Q_t \leq T | Q_0\Big) d\mu(Q_0) \\
&\overset{(c)}{=} \int_T^{T+1} p^{\pi^T}(T) e^{-\alpha_0(Q_0 - T)} d\mu(Q_0) \\
&= \mathbf{E}[e^{-\alpha_0(Q_{\tau_T} - T)} | \tau_T < \tau_e] p^{\pi^T}(T),
\end{aligned}
\tag{41}
$$

where μ denotes the conditional distribution of Q_{τ_T} given $\tau_T < \tau_e$. Note that $Q_{\tau_T} \in [T, T+1]$ because the size of the overshoot is bounded by one. Further, (a) follows from stationarity of the arrival processes and the control policy, (b) holds because a necessary condition for the interruption event is to cross the threshold T when starting from a point $Q_0 \geq T$. Finally (c) follows from Lemma 1 and the definition of the risky policy. The relations (40) and (41) together result in

$$
p^{\pi^T}(T) = \frac{e^{-\alpha_1 T}\big(1 - \mathbf{E}_\mu[e^{-\alpha_1(Q_{\tau_T} - T)}]\big)}{1 - \mathbf{E}_\mu[e^{-\alpha_0(Q_{\tau_T} - T)}] + \kappa},
\tag{42}
$$

where $\kappa = \mathbf{E}_\mu[e^{-\alpha_0 Q_{\tau_T} - (\alpha_1 - \alpha_0)T}] - \mathbf{E}_\mu[e^{-\alpha_1 Q_{\tau_T}}] \geq 0$. Therefore, using the fact that

$$
1 - x \leq e^{-x} \leq 1 - x + \frac{x^2}{2}, \quad \text{for all } x \geq 0,
\tag{43}
$$

we can provide the following bound

$$
\begin{aligned}
p^{\pi^T}(T) &\leq \frac{e^{-\alpha_1 T}\big(\alpha_1 \mathbf{E}_\mu[Q_{\tau_T} - T]\big)}{\alpha_0 \mathbf{E}_\mu[Q_{\tau_T} - T]\Big(1 - \frac{\alpha_0}{2} \cdot \frac{\mathbf{E}_\mu[(Q_{\tau_T} - T)^2]}{\mathbf{E}_\mu[Q_{\tau_T} - T]}\Big)} \\
&\leq \frac{\alpha_1}{\alpha_0(1 - \frac{\alpha_0}{2})} e^{-\alpha_1 T} \\
&= \beta e^{-\alpha_1 T},
\end{aligned}
\tag{44}
$$

where the last inequality holds, since $0 \leq Q_{\tau_{\bar{D}}} - \bar{D} \leq 1$.

Now we prove feasibility of the risky policy π^{T^*} when $D > \bar{D}$. Observe that by (13), $D > T^*$, hence the behavior of the buffer size Q_t is the same as the one in a single-server system with rate R_0 until the threshold T^* is crossed. Thus

$$
\begin{aligned}
p^{\pi^{T^*}}(D) &= \mathbf{Pr}\Big(\min_{t \geq 0} Q_t \leq 0 \big| Q_0 = D\Big) \\
&= \mathbf{Pr}\Big(\min_{t \geq 0} Q_t \leq 0 \big| \min_{t \geq 0} Q_t \leq T^*, Q_0 = D\Big) \mathbf{Pr}\Big(\min_{t \geq 0} Q_t \leq T^* | Q_0 = D\Big) \\
&= p^{\pi^{T^*}}(T^*) e^{-\alpha_0(D - T^*)} \\
&\leq \beta e^{-(\alpha_1 - \alpha_0)T^* - \alpha_0 D} = \epsilon,
\end{aligned}
$$

where the inequality follows from (44), and the last equality holds by (13).

Next we verify the feasibility of the policy π^{T^*} for $D \leq \bar{D}$. In this case, $D \leq T^*$ and by definition of the risky policy the system behaves as a single-server system with arrival rate R_1 until the threshold T^* is crossed or the buffer size hits zero (interruption). Hence, we can bound the interruption probability as follows

$$
\begin{aligned}
p^{\pi^{T^*}}(D) &= \mathbf{Pr}(\tau_e < \tau_{T^*}) \cdot 1 + \mathbf{Pr}(\tau_{T^*} < \tau_e)\mathbf{Pr}\Big(\min_{t \geq \tau_{T^*}} Q_t \leq 0 \big| \tau_{T^*} < \tau_e, Q_0 = D\Big) \\
&\overset{(a)}{=} 1 - \mathbf{Pr}(\tau_{T^*} < \tau_e)\Big(1 - \mathbf{E}_\mu[e^{-\alpha_0(Q_{\tau_{T^*}} - T^*)}] p^{\pi^{T^*}}(T^*)\Big) \\
&\overset{(b)}{=} 1 - \frac{1 - e^{-\alpha_1 D}}{1 - \mathbf{E}_\mu[e^{-\alpha_1 Q_{\tau_{T^*}}}]}\Big(1 - \mathbf{E}_\mu[e^{-\alpha_0(Q_{\tau_{T^*}} - T^*)}] p^{\pi^{T^*}}(T^*)\Big) \\
&\overset{(c)}{\leq} 1 - \frac{1 - e^{-\alpha_1 D}}{1 - \mathbf{E}_\mu[e^{-\alpha_1 Q_{\tau_{T^*}}}]}\Big(1 - \mathbf{E}_\mu[e^{-\alpha_0(Q_{\tau_{T^*}} - T^*)}] \beta e^{-\alpha_1 T^*}\Big) \\
&\overset{(d)}{\leq} \frac{(\beta - 1)(1 - e^{-\alpha_1 D})}{1 - \mathbf{E}_\mu[e^{-\alpha_1 Q_{\tau_{T^*}}}]} + 1 - \beta(1 - e^{-\alpha_1 D}) \\
&\overset{(e)}{\leq} \frac{(\beta - 1)(1 - e^{-\alpha_1 D})}{1 - e^{-\alpha_1 T^*}} + 1 - \beta(1 - e^{-\alpha_1 D}) \overset{(f)}{=} \epsilon,
\end{aligned}
$$

where (a) follows from (41), (b) is a direct consequence of Lemma 3, (c) is a result of (44), (d) may be verified by noting that $\alpha_0 = I(R_0)$, $\alpha_1 = I(R_1)$ and $R_1 \geq R_0$, (e) holds since $\beta \geq 1$ and $Q_{\tau_{T^*}} \geq T^*$. Finally, (f) immediately follows from plugging in the definition of T^* from (13).

Therefore, the risky policy π^{T^*} is feasible by Definition 2. Observe that the buffer size under any policy π^T of the form (12) with $T \geq T^*$ stochastically dominates that of policy π^{T^*}, because π^T switches to the costly server earlier, and stays in that state longer. Hence, π^T is feasible for all $T \geq T^*$. $\blacksquare$

Proof of Theorem 6. Similarly to the proof of Theorem 5, we first consider the risky policy π^T with the initial buffer size T. By definition of π^T, the costly server is used until the threshold T is crossed. Thus the expected cost of this policy is bounded by the expected time until crossing the threshold plus the expected cost given that the threshold is crossed, i.e.,

$$J^{\pi^T}(T, \epsilon) \leq \frac{\mathbf{E}[Q_{\tau_T}] - T}{R_1 - 1} + E[e^{-\alpha_0(Q_{\tau_T} - T)}] J^{\pi^T}(T, \epsilon),$$

where τ_T is defined in (38). The above relation implies

$$
\begin{aligned}
J^{\pi^T}(T, \epsilon) &\leq \frac{1}{R_1 - 1} \cdot \frac{\mathbf{E}[Q_{\tau_T} - T]}{1 - \mathbf{E}[e^{-\alpha_0(Q_{\tau_T} - T)}]} \\
&\leq \frac{1}{R_1 - 1} \cdot \frac{\mathbf{E}[Q_{\tau_T} - T]}{1 - \mathbf{E}\left[1 - \alpha_0(Q_{\tau_T} - T) + \frac{\alpha_0^2}{2}(Q_{\tau_T} - T)^2\right]} \\
&= \frac{1}{R_1 - 1} \cdot \frac{1}{\alpha_0\left(1 - \frac{\alpha_0}{2} \cdot \frac{\mathbf{E}[(Q_{\tau_T} - T)^2]}{\mathbf{E}[Q_{\tau_T} - T]}\right)} \\
&\leq \frac{1}{\alpha_0(R_1 - 1)(1 - \frac{\alpha_0}{2})} = \frac{\beta}{\alpha_1(R_1 - 1)},
\end{aligned}
\tag{45}
$$

where the second inequality follows from the fact in (43), and the last equality holds by definition of β. Now for any $D \geq \bar{D}$ we can write

$$
\begin{aligned}
J^{\pi^{T^*}}(D, \epsilon) &= \mathbf{Pr}\left(\min_{t \geq 0} Q_t \leq T^* \big| Q_0 = D\right) J^{\pi^{T^*}}(T^*, \epsilon) \\
&= e^{-a_0(D - T^*)} J^{\pi^{T^*}}(T^*, \epsilon)
\end{aligned}
$$

where the inequality holds by Lemma 1. Combining this with (45) gives the result in (14).

If $D \leq \bar{D}$, the risky policy uses the costly server until the threshold T^* is crossed at τ_{T^*} or the interruption event (τ_e), whichever happens first. Afterwards, no extra cost is incurred if an

interruption has occurred. Otherwise, by (45) an extra cost of at most $\frac{\beta}{\alpha_1(R_1-1)}$ is incurred, i.e.,

$$J^{\pi^{T^*}}(D, \epsilon) \leq \mathbf{E}\big[\min\{\tau_e, \tau_{T^*}\}\big] + \mathbf{Pr}(\tau_{T^*} < \tau_e)\frac{\beta}{\alpha_1(R_1-1)}.$$

By Doob's optional stopping theorem applied to the Martingale $Z_t = Q_t - (R_1 - 1)t$, we obtain

$$D = \mathbf{Pr}(\tau_{T^*} < \tau_e)\mathbf{E}[Q_{\tau_{T^*}}|\tau_{T^*} < \tau_e] - (R_1 - 1)\mathbf{E}\big[\min\{\tau_e, \tau_{T^*}\}\big],$$

which implies

$$\mathbf{E}\big[\min\{\tau_e, \tau_{T^*}\}\big] \leq \frac{\mathbf{Pr}(\tau_{T^*} < \tau_e)(T^* + 1) - D}{R_1 - 1}.$$

By combining the preceding relations we conclude that

$$J^{\pi^{T^*}}(D, \epsilon) \leq \frac{\mathbf{Pr}(\tau_{T^*} < \tau_e)}{R_1 - 1}\Big(T^* + 1 + \frac{\beta}{\alpha_1}\Big) - \frac{D}{R_1 - 1},$$

which immediately implies (15) by employing Lemma 3. ∎

APPENDIX B

ANALYSIS OF THE THRESHOLD POLICY FOR THE FLUID APPROXIMATION MODEL

Proof of Lemma 2. There are multiple approaches to prove the claim. We prove a more general case using Doob's optional stopping theorem that will be useful in the later arguments. We only consider $i = 0$; the other case is the same.

Let $Y_t = e^{-\theta_0 Q_t}$. It is straightforward to show that Y_t is a Martingale with respect to W_t. Now consider the boundary crossing problem, where we are interested in the probability of hitting zero before a boundary $b > D$. Let τ denote the hitting time of either boundaries. For any $n > 0$, we may apply Doob's optional stopping theorem [28] to the stopped Martingale $Y_{\tau \wedge n}$ to write

$$\mathbf{E}[Y_{\tau \wedge b}] = \mathbf{E}[e^{-\theta_0 Q_{\tau \wedge n}}] = e^{-\theta_0 D}, \quad \text{for all } n.$$

Now, we take the limit as $n \to \infty$ and exploit the dominant convergence theorem to establish:

$$\mathbf{E}[Y_\tau] = \mathbf{E}[e^{-\theta_0 Q_\tau}] = \lim_{n \to \infty} \mathbf{E}[e^{-\theta_0 Q_{\tau \wedge n}}] = e^{-\theta_0 D}. \tag{46}$$

Finally, using the Borel-Cantelli Lemma, we can show τ is finite with probability one, which allows us to decompose (46) and characterize the boundary crossing probabilities as

$$\mathbf{Pr}(Q_\tau = 0) \cdot 1 + \mathbf{Pr}(Q_\tau = b) \cdot e^{-\theta_0 b} = e^{-\theta_0 D},$$

$$\mathbf{Pr}(Q_\tau = 0) + \mathbf{Pr}(Q_\tau = b) = 1.$$

Solving the above equations gives

$$\mathbf{Pr}(Q_\tau = 0) = \frac{e^{-\theta_0 D} - e^{-\theta_0 b}}{1 - e^{-\theta_0 b}}, \tag{47}$$

$$\mathbf{Pr}(Q_\tau = b) = \frac{1 - e^{-\theta_0 D}}{1 - e^{-\theta_0 b}}. \tag{48}$$

Taking the limit as $b \to \infty$ proves the claim. ∎

Proof of Theorem 7. We first characterize the interruption probability for the cases $D > T$ and $D < T$ given $p(T)$, which is the interruption probability starting from $D = T$.

For any $x \geq 0$, define τ_x as the first hitting time of boundary x, i.e.,

$$\tau_x = \inf\{t \geq 0 : Q_t = x\}. \tag{49}$$

For the case $D > T$, using path-continuity of Q_t, strong Markov property and Lemma 2, we have

$$p(D) = \mathbf{Pr}(\tau_0 < \infty | Q_0 = D)$$

$$= \mathbf{Pr}(\tau_0 < \infty | Q_0 = T) \cdot \mathbf{Pr}(\tau_T < \infty | Q_0 = D)$$

$$= e^{-\theta_0(D-T)} p(T).$$

For the case $D < T$, we use the boundary crossing probabilities (47) and (48) that we derived in the proof of Lemma 2. Note that for the threshold policy when $D < T$, the drift is set to $R_1 - 1 = 2\theta_1$. Hence, by total probability theorem and strong Markov property, we obtain

$$p(D) = \mathbf{Pr}(\tau_0 < \infty | Q_0 = D)$$

$$= 1 \cdot \mathbf{Pr}(\tau_0 < \tau_T | Q_0 = D) + \mathbf{Pr}(\tau_0 < \infty | Q_0 = T) \cdot \mathbf{Pr}(\tau_T < \tau_0 | Q_0 = D)$$

$$= \frac{e^{-\theta_1 D} - e^{-\theta_1 b}}{1 - e^{-\theta_1 b}} + p(T) \frac{1 - e^{-\theta_1 D}}{1 - e^{-\theta_1 b}}.$$

We may obtain the desired result after simple manipulations of the above relation, once we compute $p(T)$.

In order to characterize $p(T)$, we use an analogue of one-step deviation analysis for Markov chains. Let $Q_0 = T$, and consider a small deviation Q_h, where h is a small time-step. Since Q_t is a Brownian motion with drift, Q_h has a normal distribution with variance h, and mean of $T + \alpha h$, where $\alpha \in [R_0 - 1, R_1 - 1]$. Therefore, the probability of $Q_h \geq T$ is $(\frac{1}{2} + \delta) + o(h)$, where δ is a small constant of the same order of h, and $\frac{o(h)}{h} \to 0$ as $h \to 0$. By strong Markov property of the Brownian motion, (47) and (48), we have

$$
\begin{aligned}
p(T) &= \mathbf{Pr}(\tau_0 < \infty | Q_0 = T) \\
&= \mathbf{Pr}(\tau_0 < \infty | Q_h \geq T)(\frac{1}{2} + \delta) + \mathbf{Pr}(\tau_0 < \infty | Q_h < T)(\frac{1}{2} - \delta) + o(h) \\
&= \Big[0 + p(T)\mathbf{E}_{Q_h}[\mathbf{Pr}(\tau_T < \infty) | Q_h \geq T]\Big](\frac{1}{2} + \delta) \\
&\quad + \Big[1 \cdot \mathbf{E}_{Q_h}[\mathbf{Pr}(\tau_0 < \tau_T) | Q_h < T] + p(T)\mathbf{E}_{Q_h}[\mathbf{Pr}(\tau_T < \tau_0) | Q_h < T]\Big](\frac{1}{2} - \delta) \\
&\quad + o(h) \\
&= \Big[0 + p(T)\mathbf{E}[e^{-\theta_0 Z} | Z \geq 0]\Big](\frac{1}{2} + \delta) \\
&\quad + \mathbf{E}\Big[\frac{e^{-\theta_1 T}(e^{-\theta_1 Z} - 1)}{1 - e^{-\theta_1 T}}\Big| Z < 0\Big](\frac{1}{2} - \delta) \\
&\quad + p(T)\mathbf{E}\Big[1 - \frac{e^{-\theta_1 T}(e^{-\theta_1 Z} - 1)}{1 - e^{-\theta_1 T}}\Big| Z < 0\Big](\frac{1}{2} - \delta) + o(h),
\end{aligned}
\tag{50}
$$

where $Z = Q_h - T$ is a Gaussian random variable with mean αh and variance h. In order to obtain $p(T)$, we need to compute $E[e^{-\theta_0 Z} | Z \geq 0]$ and $E[e^{-\theta_1 Z} | Z < 0]$. We may compute these expressions exactly, but it is simpler to compute upper and lower bounds and then take the limit as $h \to 0$. By (43), we have

$$
\begin{aligned}
E[e^{-\theta_0 Z} | Z \geq 0] &\leq \mathbf{E}[1 - \theta_0 Z + \frac{(\theta_0 Z)^2}{2} | Z \geq 0] = 1 - \theta_0 \beta \sqrt{h} + o(\sqrt{h}), \\
E[e^{-\theta_0 Z} | Z \geq 0] &\geq \mathbf{E}[1 - \theta_0 Z | Z \geq 0] = 1 - \theta_0 \beta \sqrt{h} + o(\sqrt{h}),
\end{aligned}
\tag{51}
$$

where β is a constant. Similarly, we get

$$
1 + \theta_1 \beta \sqrt{h} + o(\sqrt{h}) \leq E[e^{-\theta_1 Z} | Z < 0] \leq 1 + \theta_1 \beta \sqrt{h} + o(\sqrt{h}).
$$

Plugging these relations back in (50), dividing by $\beta\sqrt{h}$ and taking the limit as h goes to zero, we obtain the following equation

$$
p(T)\Big[\theta_0 + \frac{e^{-\theta_1 T}}{1 - e^{-\theta_1 T}}\theta_1\Big] = \frac{e^{-\theta_1 T}}{1 - e^{-\theta_1 T}}\theta_1,
\tag{52}
$$

which gives the desired result for $p(T)$ after rearranging the terms. ∎

Proof of Theorem 8. The proof technique for this theorem is analogous to that of Theorem 7. First, we consider the cases $D > T$ and $D < T$ and characterize the expected cost in terms of $J(T)$. Let τ_x be defined as in (49).

For the case $D > T$, note that no cost is incurred until the threshold T is reached. Hence

$$
\begin{aligned}
J(D) &= \mathbf{E}\left[\int_0^{\tau_e} u_t dt \,\middle|\, Q_0 = D\right] \\
&= \mathbf{E}\left[\int_0^{\tau_e} u_t dt \,\middle|\, \tau_T = \infty, Q_0 = D\right]\mathbf{Pr}(\tau_T = \infty | Q_0 = D) \\
&\quad + \mathbf{E}\left[\int_0^{\tau_e} u_t dt \,\middle|\, \tau_T < \infty, Q_0 = D\right]\mathbf{Pr}(\tau_T < \infty | Q_0 = D) \\
&= 0 + \mathbf{E}\left[0 + \int_{\tau_T}^{\tau_e} u_t dt \,\middle|\, \tau_T < \infty, Q_0 = D\right]\mathbf{Pr}(\tau_T < \infty | Q_0 = D) \\
&\overset{(a)}{=} \mathbf{E}\left[\int_0^{\tau_e} u_t dt \,\middle|\, Q_0 = T\right]\mathbf{Pr}(\tau_T < \infty | Q_0 = D) \\
&= J(T)\mathbf{Pr}(\tau_T < \infty | Q_0 = D) \overset{(b)}{=} J(T)e^{-\theta_0(D-T)},
\end{aligned}
$$

where (a) follows from the memoryless property of Brownian motion and (b) is a consequence of Lemma 2.

For the case $D < T$, we can use a strong Markov property to write the following for a small time-step h:

$$
\begin{aligned}
J(D) &= J(Q_0) = 1 \cdot h + \mathbf{E}_W[J(Q_h)] \\
&= h + \mathbf{E}_W\left[J(D) + \frac{\partial J}{\partial D}((R_1 - 1)h + W_h) + \frac{1}{2} \cdot \frac{\partial^2 J}{\partial D^2} \cdot h\right] + o(h) \\
&= h + J(D) + (R_1 - 1)\frac{\partial J}{\partial D} \cdot h + \frac{1}{2} \cdot \frac{\partial^2 J}{\partial D^2} \cdot h + o(h),
\end{aligned}
$$

which gives the following ordinary differential equation after dividing by h and taking the limit as $h \to 0$

$$
\frac{\partial^2 J}{\partial D^2} + \theta_1 \frac{\partial J}{\partial D} + 2 = 0, \quad 0 \le D \le T. \tag{53}
$$

It is straightforward to solve the differential equation in (53) with the boundary condition $J(0) = 0$, and $J(T)$ as a parameter. This result completes the characterization of $J(D)$ described

in (30) as a function of $J(T)$. Also, note that if we set the boundary condition $J(T) = 0$, $J(D)$ gives the expected time to hit either of the boundaries at 0 or T, i.e., we get

$$\mathbf{E}\Big[\min\{\tau_0, \tau_T\}\big|Q_0 = D\Big] = \frac{2}{\theta_1}\Big[T \cdot \frac{1 - e^{-\theta_1 D}}{1 - e^{-\theta_1 T}} - D\Big] \tag{54}$$

Now, we use a similar technique as in the proof of Theorem 7 to compute $J(T)$. Consider a small time step deviation $h > 0$ from the initial condition $Q_0 = T$. Similarly to (50), we have

$$
\begin{aligned}
J(T) &= \gamma h + \Big[J(T)\mathbf{E}_{Q_h}[\mathbf{Pr}(\tau_T < \infty)|Q_h \geq T]\Big]\mathbf{Pr}(Q_h \geq T|Q_0 = T) \\
&\quad + \Big[1 \cdot \mathbf{E}\Big[\min\{\tau_0, \tau_T\}\big|Q_h < T\Big] + 0 \cdot \mathbf{E}_{Q_h}[\mathbf{Pr}(\tau_0 < \tau_T)|Q_h < T] \\
&\quad\quad + J(T)\mathbf{E}_{Q_h}[\mathbf{Pr}(\tau_T < \tau_0)|Q_h < T]\Big]\mathbf{Pr}(Q_h < T|Q_0 = T) + o(h) \\
&= \gamma h + J(T)\mathbf{E}[e^{-\theta_0 Z}|Z \geq 0](\frac{1}{2} + \delta) \\
&\quad + \mathbf{E}\Big[\min\{\tau_0, \tau_T\}\big|Q_h < T\Big](\frac{1}{2} - \delta) \\
&\quad + J(T)\mathbf{E}\Big[1 - \frac{e^{-\theta_1 T}(e^{-\theta_1 Z} - 1)}{1 - e^{-\theta_1 T}}\Big|Z < 0\Big](\frac{1}{2} - \delta) + o(h),
\end{aligned}
$$

where γ is a constant bounded by 1, $\delta = \Theta(h)$, and $Z = Q_h - T$ is a Gaussian random variable with mean αh and variance h for some constant α. The second inequality in the preceding relations follows from (48) and Lemma 2. By (54), applying the bounds in (51) and (52), dividing by $\beta\sqrt{h}$ and taking the limit at $h \to 0$, we obtain the following equation

$$J(T)\Big[\theta_1 \frac{e^{-\theta_1 T}}{1 - e^{-\theta_1 T}} + \theta_0\Big] = \frac{2}{\theta_2}\Big[1 - \theta_1 T \frac{e^{-\theta_1 T}}{1 - e^{-\theta_1 T}}\Big],$$

which gives us the desired expression in (31) after rearranging the terms. ∎

Proof of Theorem 9. In order to facilitate verification of the HJB equation (26), we rewrite and slightly manipulate the candidate solution $V(Q, p)$ given by (32). Recall that

$$V(Q, p) = \begin{cases} V_0(Q, p), & Q \geq T(Q, p) \\ V_1(Q, p), & Q \leq T(Q, p), \end{cases} \tag{55}$$

where

$$V_0(Q, p) = e^{-\theta_0(Q - T(Q,p))}J(T(Q, p)), \tag{56}$$

$$V_1(Q, p) = \Big[J(T(Q, p)) + \frac{2}{\theta_1}T(Q, p)\Big]\frac{1 - e^{-\theta_1 Q}}{1 - e^{-\theta_1 T(Q,p)}} - \frac{2}{\theta_1}Q, \tag{57}$$

$$J(T(Q,p)) = \frac{\frac{2}{\theta_1^2}\left[1 - (1 + \theta_1 T(Q,p))e^{-\theta_1 T(Q,p)}\right]}{\frac{\theta_0}{\theta_1} + (1 - \frac{\theta_0}{\theta_1})e^{-\theta_1 T(Q,p)}}. \tag{58}$$

Note that $p^{T(Q,p)}(Q) = p$; and by definition of $p^T(\cdot)$ in (29) we may verify that the condition $Q \gtreqless T(Q,p)$ is equivalent to $p \gtreqless \frac{e^{-\theta_1 Q}}{\frac{\theta_0}{\theta_1} + (1 - \frac{\theta_0}{\theta_1})e^{-\theta_1 Q}}$. Therefore, we can partition the feasible region $\mathcal{R}$ into two sub-regions $\mathcal{R}_0$ and $\mathcal{R}_1$, such that

$$\mathcal{R}_0 = \left\{(Q,p) : p \geq \frac{e^{-\theta_1 Q}}{\frac{\theta_0}{\theta_1} + (1 - \frac{\theta_0}{\theta_1})e^{-\theta_1 Q}}\right\} \cap \mathcal{R},$$

$$\mathcal{R}_1 = \left\{(Q,p) : p < \frac{e^{-\theta_1 Q}}{\frac{\theta_0}{\theta_1} + (1 - \frac{\theta_0}{\theta_1})e^{-\theta_1 Q}}\right\} \cap \mathcal{R}.$$

Hence, we need to verify HJB for two regions separately, using the proper expression in (55).

In order to verify the HJB equation for the candidate solution (32), we also need to characterize the optimal value of the minimization problem in (26). First, we characterize the optimal solution pair $(u^*, \hat{p}^*)$ for any feasible state $(Q,p) \in \mathcal{R}$. Observe that the optimization problem in (26) can be decomposed into two smaller problems:

$$u^*(Q,p) = \operatorname{argmin}_{u \in \{0,1\}}\left\{u + \frac{\partial V}{\partial Q}(R_u - 1)\right\}, \tag{59}$$

$$\hat{p}^*(Q,p) = \operatorname{argmin}_{\hat{p}}\left\{\frac{1}{2}\frac{\partial^2 V}{\partial p^2}(\hat{p})^2 + \frac{\partial^2 V}{\partial Q \partial p}\hat{p}\right\}. \tag{60}$$

The minimization problem in (60) is quadratic and hence convex in $\hat{p}$. So we can use first order optimality condition to get

$$\hat{p}^*(Q,p) = -\frac{\partial^2 V}{\partial Q \partial p}(Q,p)/\frac{\partial^2 V}{\partial p^2}(Q,p). \tag{61}$$

For the problem in (59), $u^*(Q,p) = 0$ is and only if

$$0 + \frac{\partial V}{\partial Q}(R_0 - 1) \leq 1 + \frac{\partial V}{\partial Q}(R_1 - 1),$$

or equivalently

$$\frac{\partial V}{\partial Q}(Q,p) \geq -\frac{1}{R_1 - R_0}. \tag{62}$$

Using the chain rule and the implicit function theorem, we can analytically calculate $\frac{\partial V}{\partial Q}$ from (55) to conclude that the condition in (62) holds if and only if $(Q,p) \in \mathcal{R}_0$. In other words, $u^*(Q,p) = 0$ for all $(Q,p) \in \mathcal{R}_0$ and $u^*(Q,p) = 1$ for all $(Q,p) \in \mathcal{R}_1$.

In summary, the HJB equation in (26) boils down to the following equations:

$$0 = \frac{\theta_0}{2}\frac{\partial V_0}{\partial Q} + \frac{1}{2}\frac{\partial^2 V_0}{\partial Q^2} - \frac{1}{2}\Big(\frac{\partial^2 V_0}{\partial Q \partial p}\Big)^2 / \frac{\partial^2 V_0}{\partial p^2}, \quad \text{for all } (Q,p) \in \mathcal{R}_0, \tag{63}$$

$$0 = 1 + \frac{\theta_1}{2}\frac{\partial V_1}{\partial Q} + \frac{1}{2}\frac{\partial^2 V_1}{\partial Q^2} - \frac{1}{2}\Big(\frac{\partial^2 V_1}{\partial Q \partial p}\Big)^2 / \frac{\partial^2 V_1}{\partial p^2}, \quad \text{for all } (Q,p) \in \mathcal{R}_1, \tag{64}$$

where $V_0(Q,p)$ and $V_1(Q,p)$ are given by (56) and (57), respectively. The verification of (63) and (64) is straightforward but tedious. We omit the details for brevity. We may simply use symbolic analysis tools such as Mathematica for this part. ∎

Proof of Theorem 10. From the proof of Theorem 9, we have characterized the optimal policy $\pi^* : \mathbb{R} \times [0,1] \to \{0,1\} \times \mathbb{R}$ as

$$\pi^*(Q,p) = \Big(u^*(Q,p), \hat{p}^*(Q,p)\Big),$$

where $u^*(Q,p) = 0$ if and only if $(Q,p) \in \mathcal{R}_0$ and $\hat{p}^*(Q,p)$ given by (61) can be explicitly computed as follows:

$$\hat{p}^*(Q,p) = -\theta_0 p, \quad \text{for all } (Q,p) \in \mathcal{R}_0,$$

$$\hat{p}^*(Q,p) = -\Big[\frac{(1-p)e^{\theta_1 Q}}{(1-pe^{\theta_1 Q})^2} - \frac{1}{1-pe^{\theta_1 Q}}\Big] / \Big[\frac{(1-p)(e^{2\theta_1 Q} - e^{\theta_1 Q})}{\theta_1(1-pe^{\theta_1 Q})^2} - \frac{2(e^{\theta_1 Q}-1)}{\theta_1(1-pe^{\theta_1 Q})}\Big]$$

$$= -\frac{\theta_1}{(1+p)e^{\theta_1 Q} - 2}, \quad \text{for all } (Q,p) \in \mathcal{R}_1. \tag{65}$$

Moreover, the dynamics of the state process under the optimal control policy is given by:

$$dQ_t^* = (R_{u^*(Q_t^*, p_t^*)} - 1)dt + dW_t,$$

$$dp_t^* = \hat{p}^*(Q_t^*, p_t^*)dW_t.$$

For the proof of the first claim, observe that for a given manifold $\mathcal{M}(D, \epsilon)$, we have

$$T(Q,p) = T(D, \epsilon), \quad \text{for all } (Q,p) \in \mathcal{M}(D, \epsilon). \tag{66}$$

This claim holds by definition of $\mathcal{M}(D, \epsilon)$ in (33), and the fact that $T(Q,p)$ is the unique solution of $p^T(Q) = p$. Next, we show that if $(Q_t^*, p_t^*) \in \mathcal{M}(D, \epsilon)$ for any $t \geq 0$, then after executing the optimal policy π^*, the state process stays on the manifold $\mathcal{M}(D, \epsilon)$. First, consider the case where $Q_t^* \geq T(Q_t^*, p_t^*) = T(D, \epsilon)$. In this case, $u^*(Q_t^*, p_t^*) = 0$ and $\hat{p}^*(Q_t^*, p_t^*) = -\theta_0 p_t^*$. We would like to show that the solution of the stochastic differential equation $dp_t^* = -\theta_0 p_t^* dW_t$ coincides with the invariant manifold given by $\tilde{p}_t = e^{-\theta_0(Q_t^* - T(D, \epsilon))}p(T(D, \epsilon))$. By employing Itō's Lemma we

can check that $e^{-\theta_0(Q_t^*-T(D,\epsilon))}p(T(D,\epsilon))$ is indeed the desired solution. In particular, using the proper evolution of the queue-length process Q_t^*, we can write

$$
\begin{aligned}
d\tilde{p}_t &= -\theta_0 e^{-\theta_0(Q_t^*-T(D,\epsilon))}p(T(D,\epsilon))(dQ_t^*) \\
&\quad + \frac{1}{2}\theta_0^2 e^{-\theta_0(Q_t^*-T(D,\epsilon))}p(T(D,\epsilon))dt \\
&= e^{-\theta_0(Q_t^*-T(D,\epsilon))}p(T(D,\epsilon))\Big[-\theta_0(\frac{\theta_0}{2}dt+dW_t)+\frac{1}{2}\theta_0^2\Big] \\
&= -\theta_0 e^{-\theta_0(Q_t^*-T(D,\epsilon))}p(T(D,\epsilon))dW_t \\
&= -p_t^* dW_t.
\end{aligned}
$$

Next, consider the case where $Q_t^* < T(Q_t^*,p_t^*) = T(D,\epsilon)$. In this case, $u^*(Q_t^*,p_t^*) = 1$ and $\hat{p}^*(Q_t^*,p_t^*)$ is given by (65). Similarly to the previous case, we may use Itō's Lemma to verify that the state process stays on the invariant manifold given by

$$
\tilde{p}_t = p^{T(D,\epsilon)}(Q_t^*) = e^{-\theta_1 Q_t^*} + p(T(D,\epsilon))(1-\frac{\theta_0}{\theta_1})\Big(1-e^{-\theta_1 Q_t^*}\Big).
$$

By Itō's Lemma we have

$$
\begin{aligned}
d\tilde{p}_t &= \frac{\partial p^{T(D,\epsilon)}(Q_t^*)}{\partial Q}dQ_t^* + \frac{1}{2}\cdot\frac{\partial^2 p^{T(D,\epsilon)}(Q_t^*)}{\partial Q^2}dt \\
&= -\theta_1 e^{-\theta_1 Q_t^*}\Big[1-p(T(D,\epsilon))(1-\frac{\theta_0}{\theta_1})\Big](\frac{\theta_1}{2}dt+dW_t) \\
&\quad +\frac{1}{2}\cdot\theta_1^2 e^{-\theta_1 Q_t^*}\Big[1-p(T(D,\epsilon))(1-\frac{\theta_0}{\theta_1})\Big]dt \\
&= -p_t^* dW_t = dp_t^*,
\end{aligned}
$$

which completes the proof of the first claim.

Now that we have established that the state process starting from (D,ϵ) under optimal control stays on a one-dimensional invariant manifold $\mathcal{M}(D,\epsilon)$, the optimality of the threshold policy $\pi^{T(D,\epsilon)}(Q)$ is immediate. Recall that the decision process of importance is $u_t^* \in \{0,1\}$, and we know that $u^*(Q,p) = 0$ if and only if $Q \geq T(Q,p)$. Moreover, since the optimal state process stays on $\mathcal{M}(D,\epsilon)$, we have $T(Q_t^*,p_t^*) = T(D,\epsilon)$. Hence, the optimal control policy (given the initial condition) chooses the action $u^*(Q,p) = 0$ if and only if $Q \geq T(D,\epsilon)$. Therefore, the optimal policy $\pi^*(Q,p)$ coincides with the threshold policy $\pi^{T(D,\epsilon)}(Q)$. We may also verify that the interruption probability under the threshold policy conditioned on the history up to time t is given by p_t^*. ∎

REFERENCES

[1] C. Labovitz, D. McPherson, and S. Iekel-Johnson. 2009 Internet Observatory report. In *NANOG-47*, October 2009.

[2] Cisco Visual Networking Index: Forecast and methodology, 2010-2015. Cisco, 2011.

[3] Cisco Visual Networking Index: Global mobile data traffic forecast update, 2010-2015. Cisco, 2011.

[4] T. Ho, R. Koetter, M. Médard, M. Effros, J. Shi, and D. Karger. A random linear network coding approach to multicast. *IEEE Transactions on Information Theory*, 52:4413–4430, 2006.

[5] A. ParandehGheibi, M. Médard, A. E. Ozdaglar, and S. Shakkottai. Avoiding interruptions - a QoE reliability function for streaming media applications. *IEEE Journal on Selected Areas in Communications*, 29(5):1064–1074, 2011.

[6] R.C. Chen. Constrained stochastic control with probabilistic criteria and search optimization. In *Proc. 43rd IEEE Conference on Decision and Control*, December 2004.

[7] M. J. Neely. Super-fast delay tradeoffs for utility optimal fair scheduling in wireless networks. *IEEE Journal on Selected Areas in Communications (JSAC), Special Issue on Nonlinear Optimization of Communication Systems*, 24(8):1489–1501, 2006.

[8] M. J. Neely. Delay-based network utility maximization. In *Proc. of INFOCOM*, San Diego, CA, March 2010.

[9] S. Lu, V. Bharghavan, and R. Srikant. Fair scheduling in wireless packet networks. *IEEE/ACM Transactions on Networking (TON)*, 7(4):473–489, 1999.

[10] R.A. Berry and R.G. Gallager. Communication over fading channels with delay constraints. *IEEE Transactions on Information Theory*, 48(5):1135–1149, 2003.

[11] I. Hou and P. R. Kumar. Scheduling heterogeneous real-time traffic over fading wireless channels. In *Proc. of INFOCOM*, San Diego, CA, 2010.

[12] Y. Zhou, D. Chiu, and J. Lui. A simple model for analyzing P2P streaming protocols. In *Proc. IEEE ICNP 2007*.

[13] T. Bonald, L. Massoulié, F. Mathieu, D. Perino, and A. Twigg. Epidemic live streaming: optimal performance trade-offs. *SIGMETRICS Perform. Eval. Rev.*, 36(1):325–336, 2008.

[14] Bridge Q. Zhao, John C.S. Lui, and Dah-Ming Chiu. Exploring the optimal chunk selection policy for data-driven P2P streaming systems. In *The 9th International Conference on Peer-to-Peer Computing*, 2009.

[15] L. Ying, R. Srikant, and S. Shakkottai. The Asymptotic Behavior of Minimum Buffer Size Requirements in Large P2P Streaming Networks. In *Proc. of the Information Theory and Applications Workshop*, San Diego, CA, 2010.

[16] D. Kumar, E. Altman, and J-M. Kelif. Globally Optimal User-Network Association in an 802.11 WLAN and 3G UMTS Hybrid Cell. In *Proc. of the 20th International Teletraffic Congress (ITC-20)*, Ottawa, 2007.

[17] E. Altman. *Constrained Markov Decision Processes*. Chapman and Hall, 1999.

[18] A.B. Piunovskiy. *Optimal Control of Random Sequences in Problems with Constraints*. Kluwer Academic Publishers, 1997.

[19] A.B. Piunovskiy. Controlled random sequences: the convex analytic approach and constrained problems. *Russian Mathematical Surveys*, 53:1233–1293, 1998.

[20] E.A. Feinberg and A. Shwartz. Constrained Markov decision models with discounted rewards. *Mathematics of Operations Research*, 20:302–320, 1995.

[21] G.L. Blankenship R.C. Chen. Dynamic programming equations for discounted constrained stochastic control. *IEEE Transactions on Automatic Control*, 49:699–709, 2004.

[22] A.B. Piunovskiy and X. Mao. Constrained Markov decision processes: the dynamic programming approach. *Operations Research Letters*, 27:119–126, 2000.

[23] R. Brockett. Stochastic control. Lecture Notes, Harvard University, Cambridge, MA, 2009.

[24] D. Wischik, M. Handley, and C. Raiciu. Control of multipath TCP and optimization of multipath routing in the internet. *Network Control and Optimization*, 5894:204–218, 2009.

[25] D. Wischik, M. Handley, and C. Raiciu. Control of multipath TCP and optimization of multipath routing in the internet. *Network Control and Optimization*, 5894:204–218, 2009.

[26] A. Ford, C. Raiciu, M. Handley, S. Barre, and J. Iyengar. Architectural guidelines for multipath TCP development. July 2011. draft-ietf-mptcp-architecture-05.

[27] Jay Kumar Sundararajan, Szymon Jakubczak, Muriel Médard, Michael Mitzenmacher, and João Barros. Interfacing network coding with TCP: an implementation. *CoRR*, abs/0908.1564, 2009.

[28] I. Karatzas and S. Shreve. *Brownian Motion and Stochastic Calculus*. Springer, 1997.

[29] T. Kurtz. Strong approximation theorems for density dependent Markov chains. *Stochastic Processes and Their Applications*, 6:223–240, 1978.